WEST

Help | Logout | Interrupt

Main Menu | Search Form | Posting Counts | Show S Numbers | Edit S Numbers | Preferences | Cases

Search Results -

Terms	Documents
19 and 15	0

Database:
US Patents Full-Text Database
US Pre-Grant Publication Full-Text Database
JPO Abstracts Database
EPO Abstracts Database
Derwent World Patents Index
IBM Technical Disclosure Bulletins

Search: L13

Refine Search

Recall Text | Clear

Search History

DATE: Thursday, April 24, 2003 Printable Copy Create Case

Set Name side by side	Query	Hit Count	Set Name result set
DB=USPT,PGPB; PLUR=YES; OP=AND			
L13	l9 and l5	0	L13
L12	l9 and l4	0	L12
L11	l9 and l3	0	L11
L10	L9 and l1	0	L10
L9	monophosphoryl adj lipid adj a or aluminum adj salt or qs21 or isa adj 720 or saf or iscoms or mf-59 or sbas-2 or sbas-4 or detox or rc-529 or glucosaminide-4-phosphate	15240	L9
L8	adjuvant and l3	0	L8
L7	adjuvant and l4	0	L7
L6	adjuvant and L5	0	L6
L5	5773246.pn.	1	L5
L4	5444149.pn.	1	L4
L3	5525495.pn.	1	L3
L2	adjuvant and L1	0	L2
L1	6255055.pn.	1	L1

END OF SEARCH HISTORY

```
=> d his

     (FILE 'HOME' ENTERED AT 15:48:12 ON 24 APR 2003)

     FILE 'MEDLINE, CAPLUS, BIOSIS, SCISEARCH' ENTERED AT 15:48:26 ON 24 APR
     2003
L1         1253 S KOC
L2            0 S MUELLER-PALLASCH AND L1
L3           71 S L1 AND CANCER
L4           36 DUP REM L3 (35 DUPLICATES REMOVED)

=> d au ti so 1-36 l4

L4   ANSWER 1 OF 36     MEDLINE                          DUPLICATE 1
AU   Wang T; Fan L; Watanabe Y; McNeill P D; Moulton G G; Bangur C; Fanger G R;
     Okada M; Inoue Y; Persing D H; Reed S G
TI   L523S, an RNA-binding protein as a potential therapeutic target for lung
     cancer.
SO   BRITISH JOURNAL OF CANCER, (2003 Mar 24) 88 (6) 887-94.
     Journal code: 0370635. ISSN: 0007-0920.

L4   ANSWER 2 OF 36     MEDLINE                          DUPLICATE 2
AU   Mueller F; Bommer M; Lacher U; Ruhland C; Stagge V; Adler G; Gress T M;
     Seufferlein T
TI   KOC is a novel molecular indicator of malignancy.
SO   BRITISH JOURNAL OF CANCER, (2003 Mar 10) 88 (5) 699-701.
     Journal code: 0370635. ISSN: 0007-0920.

L4   ANSWER 3 OF 36     MEDLINE                          DUPLICATE 3
AU   Zhang Jian-Ying; Casiano Carlos A; Peng Xuan-Xian; Koziol James A; Chan
     Edward K L; Tan Eng M
TI   Enhancement of Antibody Detection in Cancer Using Panel of
     Recombinant Tumor-associated Antigens.
SO   CANCER EPIDEMIOLOGY, BIOMARKERS AND PREVENTION, (2003 Feb) 12 (2) 136-43.
     Journal code: 9200608. ISSN: 1055-9965.

L4   ANSWER 4 OF 36  CAPLUS  COPYRIGHT 2003 ACS        DUPLICATE 4
IN   Mack, David H.; Gish, Kurt C.
TI   Gene expression profiles useful for diagnosis of human ovarian cancer and
     screening for modulators of ovarian cancer
SO   PCT Int. Appl., 332 pp.
     CODEN: PIXXD2

L4   ANSWER 5 OF 36  CAPLUS  COPYRIGHT 2003 ACS
IN   Aziz, Natasha; Murray, Richard
TI   Gene expression profile in human lung cancer and its use in diagnosis and
     screening for modulators of lung cancer
SO   PCT Int. Appl., 453 pp.
     CODEN: PIXXD2

L4   ANSWER 6 OF 36  CAPLUS  COPYRIGHT 2003 ACS
IN   Murray, Richard; Glynne, Richard; Watson, Susan R.; Aziz, Natasha
TI   Gene expression profiles associated with angiogenesis and their use in
     diagnosis and screening for angiogenesis modulators
SO   PCT Int. Appl., 291 pp.
     CODEN: PIXXD2

L4   ANSWER 7 OF 36  CAPLUS  COPYRIGHT 2003 ACS
IN   Mueller, Friederike; Gress, Thomas; Adler, Guido
TI   Medicament comprising a DNA sequence, which codes for the RNA-binding
     KOC protein, and comprising a KOC protein or a DNA
     sequence of the KOC promoter
SO   PCT Int. Appl., 75 pp.
```

CODEN: PIXXD2

L4 ANSWER 8 OF 36 MEDLINE DUPLICATE 5
AU Hoo Linda Soo; Zhang Jianying Y; Chan Edward K L
TI Cloning and characterization of a novel 90 kDa 'companion' auto-antigen of p62 overexpressed in **canc r**.
SO ONCOGENE, (2002 Jul 25) 21 (32) 5006-15.
Journal code: 8711562. ISSN: 0950-9232.

L4 ANSWER 9 OF 36 SCISEARCH COPYRIGHT 2003 ISI (R)
AU Itamochi H (Reprint); Kigawa J; Sultana H; Iba T; Akeshima R; Kamazawa S; Kanamori Y; Terakawa N
TI Sensitivity to anticancer agents and resistance mechanisms in clear cell carcinoma of the ovary
SO JAPANESE JOURNAL OF CANCER RESEARCH, (JUN 2002) Vol. 93, No. 6, pp. 723-728.
Publisher: BUSINESS CENTER ACADEMIC SOCIETIES JAPAN, 5-16-9 HONKOMAGOME, BUNKYO-KU, TOKYO, 113-8633, JAPAN.
ISSN: 0910-5050.

L4 ANSWER 10 OF 36 CAPLUS COPYRIGHT 2003 ACS
AU Wang, Shanying; Liu, Long; Wang, Lin; Lian, Qizhou; Li, Qifu; Peng, Xuanxian
TI Auto-antibody to **Koc** in patients with **cancers**
SO Zhongguo Mianyixue Zazhi (2002), 18(1), 30-32
CODEN: ZMZAEE; ISSN: 1000-484X

L4 ANSWER 11 OF 36 MEDLINE DUPLICATE 6
AU Kiyozuka Y; Nakagawa H; Senzaki H; Uemura Y; Adachi S; Teramoto Y; Matsuyama T; Bessho K; Tsubura A
TI Bone morphogenetic protein-2 and type IV collagen expression in psammoma body forming ovarian **cancer**.
SO ANTICANCER RESEARCH, (2001 May-Jun) 21 (3B) 1723-30.
Journal code: 8102988. ISSN: 0250-7005.

L4 ANSWER 12 OF 36 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC.
AU Tan, Eng M. (1)
TI Autoantibodies as reporters identifying aberrant cellular mechanisms in tumorigenesis.
SO Journal of Clinical Investigation, (November, 2001) Vol. 108, No. 10, pp. 1411-1415. print.
ISSN: 0021-9738.

L4 ANSWER 13 OF 36 CAPLUS COPYRIGHT 2003 ACS
AU Martinez, Eduardo J.; Corey, E. J.; Owa, Takashi
TI Antitumor activity- and gene expression-based profiling of ecteinascidin Et 743 and phthalascidin Pt 650
SO Chemistry & Biology (2001), 8(12), 1151-1160
CODEN: CBOLE2; ISSN: 1074-5521

L4 ANSWER 14 OF 36 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC.
AU Kohno, T. (1); Suzuki, M. (1); Mizukami, H.; Saga, Y. (1); Takei, Y. (1); Shimpo, M.; Ohwada, M. (1); Kume, A.; Ozawa, K. (1); Sato, I. (1)
TI Expression of interleukin-10 inhibits angiogenesis and tumor growth in ovarian **cancer**.
SO European Journal of Cancer, (October, 2001) Vol. 37, No. Supplement 6, pp. S325. http://www.elsevier.com/locate/ejca. print.
Meeting Info.: 11th European Cancer Conference Lisbon, Portugal October 21-25, 2001
ISSN: 0959-8049.

L4 ANSWER 15 OF 36 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC.
AU Bommer, Martin (1); Seufferlein, Thomas; Gress, Thomas; Adler, Guido; Doehner, Hartmut (1); Mueller, Friederike

TI Detection of CNS involvement in patients with acute leukemia and Non-Hodgkin's Lymphoma using RT-PCR for **KOC**-expression.
SO Blood, (November 16, 2001) Vol. 98, No. 11 Part 2, pp. 157b. http://www.bloodjournal.org/. print.
Meeting Info.: 43rd Annual Meeting of the American Society of Hematology, Part 2 Orlando, Florida, USA December 07-11, 2001
ISSN: 0006-4971.

L4 ANSWER 16 OF 36 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC.
AU Zhang, Jianying (1); Chan, Edward K. L.; Lu, Maolong; Wang, Xiao; Peng, Xuanxian; Mueller-Pillasch, Friederike; Tan, Eng M.
TI Autoimmune responses to mRNA binding proteins p62 and **Koc** in **cancer**.
SO Proceedings of the American Association for Cancer Research Annual Meeting, (March, 2001) Vol. 42, pp. 157. print.
Meeting Info.: 92nd Annual Meeting of the American Association for Cancer Research New Orleans, LA, USA March 24-28, 2001
ISSN: 0197-016X.

L4 ANSWER 17 OF 36 MEDLINE DUPLICATE 7
AU Zhang J Y; Chan E K; Peng X X; Lu M; Wang X; Mueller F; Tan E M
TI Autoimmune responses to mRNA binding proteins p62 and **Koc** in diverse malignancies.
SO CLINICAL IMMUNOLOGY, (2001 Aug) 100 (2) 149-56.
Journal code: 100883537. ISSN: 1521-6616.

L4 ANSWER 18 OF 36 CAPLUS COPYRIGHT 2003 ACS
IN Wang, Tongtong; Fan, Liqun
TI Compounds and methods for therapy and diagnosis of lung **cancer**
SO PCT Int. Appl., 261 pp.
CODEN: PIXXD2

L4 ANSWER 19 OF 36 CAPLUS COPYRIGHT 2003 ACS DUPLICATE 8
AU Wang, Tongtong; Hopkins, Deborah; Schmidt, Cheryl; Silva, Sandra; Houghton, Raymond; Takita, Hiroshi; Repasky, Elizabeth; Reed, Steven G.
TI Identification of genes differentially over-expressed in lung squamous cell carcinoma using combination of cDNA subtraction and microarray analysis
SO Oncogene (2000), 19(12), 1519-1528
CODEN: ONCNES; ISSN: 0950-9232

L4 ANSWER 20 OF 36 MEDLINE DUPLICATE 9
AU Gure A O; Stockert E; Arden K C; Boyer A D; Viars C S; Scanlan M J; Old L J; Chen Y T
TI CT10: a new **cancer**-testis (CT) antigen homologous to CT7 and the MAGE family, identified by representational-difference analysis.
SO INTERNATIONAL JOURNAL OF CANCER, (2000 Mar 1) 85 (5) 726-32.
Journal code: 0042124. ISSN: 0020-7136.

L4 ANSWER 21 OF 36 CAPLUS COPYRIGHT 2003 ACS
IN Chen, Yao-Tseng; Gure, Ali; Tsang, Solam; Stockert, Elisabeth; Jager, Elke; Knuth, Alexander; Old, Lloyd J.
TI Isolated nucleic acid molecule encoding the **cancer** associated antigens CT7, **KOC**-2 and **KOC**-3, expression of the antigens and application to **cancer** diagnosis and immunotherapy
SO PCT Int. Appl., 45 pp.
CODEN: PIXXD2

L4 ANSWER 22 OF 36 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC.
AU Wallrapp, C.; Mueller-Pillasch, F.; Micha, A.; Wenger, C.; Geng, M.; Solinas-Toldo, S.; Lichter, P.; Frohme, M.; Hoheisel, J. D.; Adler, G.; Gress, T. M. (1)
TI Strategies for the detection of disease genes in pancreatic **cancer**
.

SO Lohr, J.-Matthias; Colcher, David; Hollingsworth, Michael A.; Liebe, Stefan. Annals of the New York Academy of Sciences, (June 30, 1999) Vol. 880, pp. 122-146. Annals of the New York Academy of Sciences; Cell and molecular biology of pancreatic carcinoma: Recent developments in research and experimental therapy. print.
Publisher: New York Academy of Sciences 2 East 63rd Street, New York, NY, 10021, USA.
Meeting Info.: Baltic Pancreas Meeting on Pancreatic Carcinoma from Bench to Bedside Warnemuende, Germany September 17-18, 1998
ISSN: 0077-8923. ISBN: 1-57331-219-3 (cloth), 1-57331-220-7 (paper).

L4 ANSWER 23 OF 36 MEDLINE DUPLICATE 10
AU Zhang J Y; Chan E K; Peng X X; Tan E M
TI A novel cytoplasmic protein with RNA-binding motifs is an autoantigen in human hepatocellular carcinoma.
SO JOURNAL OF EXPERIMENTAL MEDICINE, (1999 Apr 5) 189 (7) 1101-10.
Journal code: 2985109R. ISSN: 0022-1007.

L4 ANSWER 24 OF 36 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC.
AU Yamada, Akira (1); Kawano, Koichiro; Harashima, Nanae; Niiya, Fumihiko; Nagai, Kouji; Kobayashi, Terutada; Mine, Takashi; Ushijima, Kimio; Nishida, Takashi; Itoh, Kyogo
TI Study of HLA class I restriction and the directed antigens of cytotoxic T lymphocytes at the tumor sites of ovarian **cancer**.
SO Cancer Immunology Immunotherapy, (May-June, 1999) Vol. 48, No. 2-3, pp. 147-152.
ISSN: 0340-7004.

L4 ANSWER 25 OF 36 MEDLINE DUPLICATE 11
AU Mueller-Pillasch F; Pohl B; Wilda M; Lacher U; Beil M; Wallrapp C; Hameister H; Knochel W; Adler G; Gress T M
TI Expression of the highly conserved RNA binding protein **KOC** in embryogenesis.
SO MECHANISMS OF DEVELOPMENT, (1999 Oct) 88 (1) 95-9.
Journal code: 9101218. ISSN: 0925-4773.

L4 ANSWER 26 OF 36 CAPLUS COPYRIGHT 2003 ACS
AU Chen, Yao-Tseng; Gure, Ali O.; Tsang, Solam; Stockert, Elisabeth; Jager, Elke; Knuth, Alexander; Old, Lloyd J.
TI Identification of multiple **cancer**/testis antigens by allogeneic antibody screening of a melanoma cell line library
SO Proceedings of the National Academy of Sciences of the United States of America (1998), 95(12), 6919-6923
CODEN: PNASA6; ISSN: 0027-8424

L4 ANSWER 27 OF 36 CAPLUS COPYRIGHT 2003 ACS
AU Hirai, Nobuyuki
TI Establishment and characterization of ovarian yolk sac tumor cell line in nude mice
SO Kurume Igakkai Zasshi (1998), 61(12), 334-343
CODEN: KIZAAL; ISSN: 0368-5810

L4 ANSWER 28 OF 36 CAPLUS COPYRIGHT 2003 ACS
AU Tomioka, Yoshito
TI Establishment and characterization of three human ovarian clear cell carcinoma cell line
SO Kurume Igakkai Zasshi (1998), 61(12), 323-333
CODEN: KIZAAL; ISSN: 0368-5810

L4 ANSWER 29 OF 36 MEDLINE DUPLICATE 12
AU Motomura S
TI Induction of apoptosis in ovarian carcinoma cell line by gonadotropin-releasing hormone agonist.
SO KURUME MEDICAL JOURNAL, (1998) 45 (1) 27-32.

Journal code: 2985210R. ISSN: 0023-5679.

L4 ANSWER 30 OF 36 MEDLINE DUPLICATE 13
AU Mueller-Pillasch F; Lacher U; Wallrapp C; Micha A; Zimmerhackl F; Hameister H; Varga G; Friess H; Buchler M; Beger H G; Vila M R; Adler G; Gress T M
TI Cloning of a gene highly overexpressed in **cancer** coding for a novel KH-domain containing protein.
SO ONCOGENE, (1997 Jun 5) 14 (22) 2729-33.
Journal code: 8711562. ISSN: 0950-9232.

L4 ANSWER 31 OF 36 CAPLUS COPYRIGHT 2003 ACS DUPLICATE 14
AU Kataoka, Akio; Nishida, Takashi; Hirai, Nobuyuki; Tomioka, Yoshito; Sugiyama, Toru; Yakushiji, Michiaki
TI Induction of apoptosis in ovarian carcinoma cell line by glucocorticoids, and sex steroid hormones
SO Oncology Reports (1997), 4(6), 1249-1253
CODEN: OCRPEW; ISSN: 1021-335X

L4 ANSWER 32 OF 36 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC.
AU Mueller-Pillasch, Friederike (1); Lacher, Ulrike (1); Wallrapp, Christine (1); Micha, Anne (1); Zimmerhackl, Frank (1); Varga, G.; Friess, H.; Buechler, M.; Vila, M. R.; Hameister, H.; Adler, Guido (1); Gress, Thomas M. (1)
TI Cloning of a gene highly overexpressed in **cancer** coding for a novel KH-domain containing protein.
SO Gastroenterology, (1997) Vol. 112, No. 4 SUPPL., pp. A621.
Meeting Info.: Digestive Disease Week and the 97th Annual Meeting of the American Gastroenterological Association Washington, D.C., USA May 11-14, 1997
ISSN: 0016-5085.

L4 ANSWER 33 OF 36 SCISEARCH COPYRIGHT 2003 ISI (R)
AU Okina H; Kataoka A (Reprint); Sugiyama T; Nishida T; Yakushiji M
TI Establishment and characterization of ovarian endometrioid adenocarcinoma cell line in nude mice and analyses of the immunohistochemical property among the original, recurred, and heterotransplanted tumor
SO INTERNATIONAL JOURNAL OF ONCOLOGY, (FEB 1997) Vol. 10, No. 2, pp. 311-316.
Publisher: INT JOURNAL ONCOLOGY, C/O PROFESSOR D A SPANDIDOS, EDITORIAL OFFICE, 1, S MERKOURI ST, ATHENS 116 35, GREECE.
ISSN: 1019-6439.

L4 ANSWER 34 OF 36 MEDLINE DUPLICATE 15
AU Kataoka A; Nishida T; Motomura S; Higashijima T; Ueyama T; Sugiyama T; Yakushiji M
TI The effects of sex steroid hormones, gonadotropins, and gonadotropin releasing hormone agonist on the proliferation of ovarian **cancer** cell line (**KOC**-2S).
SO NIPPON SANKA FUJINKA GAKKAI ZASSHI. ACTA OBSTETRICA ET GYNAECOLOGICA JAPONICA, (1994 Oct) 46 (10) 1033-40.
Journal code: 7505749. ISSN: 0300-9165.

L4 ANSWER 35 OF 36 CAPLUS COPYRIGHT 2003 ACS
AU Akizuki, Hideaki
TI An experimental study on optimal administration of cisplatin for ovarian **cancer**
SO Kurume Igakkai Zasshi (1991), 54(5), 359-66
CODEN: KIZAAL; ISSN: 0368-5810

L4 ANSWER 36 OF 36 MEDLINE DUPLICATE 16
AU Kataoka A; Yokota D; Sugiyama T; Oda T; Nishida T; Yakushiji M
TI Effects of anticancer agents on 7, 12-dimethylbenz (a) anthracene induced rat ovarian **cancer** cell line (DMBA-OC-1) and human ovarian serous adenocarcinoma cell line (**KOC**-1S).

SO NIPPON SANKA FUJINKA GAKKAI ZASSHI. ACTA OBSTETRICA ET GYNAECOLOGICA JAPONICA, (1989 Oct) 41 (10) 1523-9.
Journal code: 7505749. ISSN: 0300-9165.

=> s (immune(w)response or t(w)cell) and l4
3 FILES SEARCHED...
L5 8 (IMMUNE(W) RESPONSE OR T(W) CELL) AND L4

=> d his

(FILE 'HOME' ENTERED AT 15:48:12 ON 24 APR 2003)

FILE 'MEDLINE, CAPLUS, BIOSIS, SCISEARCH' ENTERED AT 15:48:26 ON 24 APR 2003

L1 1253 S KOC
L2 0 S MUELLER-PALLASCH AND L1
L3 71 S L1 AND CANCER
L4 36 DUP REM L3 (35 DUPLICATES REMOVED)
L5 8 S (IMMUNE(W)RESPONSE OR T(W)CELL) AND L4

=> d au ti so ab 1-8 l5

L5 ANSWER 1 OF 8 MEDLINE
AU Wang T; Fan L; Watanabe Y; McNeill P D; Moulton G G; Bangur C; Fanger G R; Okada M; Inoue Y; Persing D H; Reed S G
TI L523S, an RNA-binding protein as a potential therapeutic target for lung **cancer**.
SO BRITISH JOURNAL OF CANCER, (2003 Mar 24) 88 (6) 887-94.
Journal code: 0370635. ISSN: 0007-0920.
AB Approaches to vaccine-based immunotherapy of human **cancer** may ultimately require targets that are both tumour-specific and immunogenic. In order to generate specific antitumour **immune responses** to lung **cancer**, we have sought lung **cancer**-specific proteins that can be targeted for adjuvant vaccine therapy. By using a combination of cDNA subtraction and microarray analysis, we previously reported the identification of an RNA-binding protein within the **KOC** family, L523S, to be overexpressed in squamous cell **cancers** of the lung. We show here that L523S exhibits significant potential for vaccine immunotherapy of lung **cancer**. As an oncofetal protein, L523S is normally expressed in early embryonic tissues, yet it is re-expressed in a high percentage of nonsmall cell lung carcinoma. The specificity of L523S expression in lung **cancer** was demonstrated by both mRNA and protein measurements using real-time PCR, Western blot, and immunohistochemistry analyses. Furthermore, we show that immunological tolerance of L523S is naturally broken in lung **cancer** patients, as evidenced by detectable antibody responses to recombinant L523S protein in eight of 17 lung pleural effusions from lung **cancer** patients. Collectively, our studies suggest that L523S may be an important marker of malignant progression in human lung **cancer**, and further suggest that treatment approaches based on L523S as an immunogenic target are worthy of pursuit.British Journal of **Cancer** (2003) 88, 887-894.
doi:10.1038/sj.bjc.6600806 www.bjcancer.com

L5 ANSWER 2 OF 8 MEDLINE
AU Zhang J Y; Chan E K; Peng X X; Lu M; Wang X; Mueller F; Tan E M
TI Autoimmune responses to mRNA binding proteins p62 and **Koc** in diverse malignancies.
SO CLINICAL IMMUNOLOGY, (2001 Aug) 100 (2) 149-56.
Journal code: 100883537. ISSN: 1521-6616.
AB Two tumor-associated antigens, p62 and **Koc**, are insulin-like growth factor II (IGF-II) messenger RNA binding proteins. Autoantibodies to p62 have been detected in **cancer** sera but have not been

reported for **Koc**. This study determined the extent and frequency of autoantibodies to p62 and **Koc** in diverse malignancies, the epitopes on the antigens, and the presence or absence of cross-reactive antibodies. Recombinant polypeptides were expressed from full-length and partial cDNA constructs and used as antigens in Western blotting, enzyme-linked immunoassay, and immunoprecipitation. After identifying the epitopes, cross-absorption with recombinant polypeptides was used to determine specificity. Sera from 777 patients with 10 different types of malignancy were analyzed. Autoantibodies to p62 were found in 11.6% and to **Koc** in 12.2% and cumulatively to both antigens in 20.5%, with significant difference from the control populations consisting of normal subjects and autoimmune disease patients ($P < 0.01$). The immunodominant epitopes were at the amino termini of both antigens and absorption studies showed that the majority of autoantibodies were not cross-reactive. Autoantibodies to p62 and **Koc** were present in approximately similar frequencies in a variety of malignancies and the **immune responses** appeared to be independent of each other. The **immune responses** might be related to overexpression or dysregulation of p62 and **Koc** in some tumors.

L5 ANSWER 3 OF 8 MEDLINE
AU Gure A O; Stockert E; Arden K C; Boyer A D; Viars C S; Scanlan M J; Old L J; Chen Y T
TI CT10: a new **cancer**-testis (CT) antigen homologous to CT7 and the MAGE family, identified by representational-difference analysis.
SO INTERNATIONAL JOURNAL OF CANCER, (2000 Mar 1) 85 (5) 726-32.
Journal code: 0042124. ISSN: 0020-7136.
AB Assays relying on humoral or **T-cell**-based recognition of tumor antigens to identify potential targets for immunotherapy have led to the discovery of a significant number of immunogenic gene products, including **cancer**-testis (CT) antigens predominantly expressed in **cancer** cells and male germ cells. The search for **cancer** -specific antigens has been extended via the technique of representational-difference analysis and SK-MEL-37, a melanoma cell line expressing a broad range of CT antigens. Using this approach, we have isolated CT antigen genes, genes over-expressed in **cancer**, e. g., PRAME and **KOC**, and genes encoding neuro-ectodermal markers. The identified CT antigen genes include the previously defined MAGE-A6, MAGE-A4a, MAGE-A10, CT7/MAGE-C1, as well as a novel gene designated CT10, which shows strong homology to CT7/MAGE-C1 both at cDNA and at genomic levels. Chromosome mapping localized CT10 to Xq27, in close proximity to CT7/MAGE-C1 and MAGE-A genes. CT10 mRNA is expressed in testis and in 20 to 30% of various human **cancers**. A serological survey identified 2 melanoma patients with anti-CT10 antibody, demonstrating the immunogenicity of CT10 in humans.

L5 ANSWER 4 OF 8 MEDLINE
AU Zhang J Y; Chan E K; Peng X X; Tan E M
TI A novel cytoplasmic protein with RNA-binding motifs is an autoantigen in human hepatocellular carcinoma.
SO JOURNAL OF EXPERIMENTAL MEDICINE, (1999 Apr 5) 189 (7) 1101-10.
Journal code: 2985109R. ISSN: 0022-1007.
AB In hepatocellular carcinoma (HCC), autoantibodies to intracellular antigens are detected in 30-40% of patients. Patients with chronic hepatitis or liver cirrhosis develop HCC, and when this occurs, some patients exhibit autoantibodies of new specificities. It has been suggested that these novel autoantibody responses may be immune system reactions to proteins involved in transformation-associated cellular events. One HCC serum shown to contain antibodies to unidentified cellular antigens was used to immunoscreen a cDNA expression library, and a full length cDNA clone was isolated with an open reading frame encoding 556 amino acids with a predicted molecular mass of 62 kD. The 62-kD

protein contained two types of RNA-binding motifs, the consensus sequence RNA-binding domain (CS-RBD) and four hnRNP K homology (KH) domains. This protein, provisionally called p62, has close identity (66-70%) to three other proteins at the amino acid sequence level, and all four proteins may belong to a family having CS-RBD in the NH2-terminal region and four KH domains in the mid-to-COOH- terminal region. The homologous proteins are: KH domain-containing protein overexpressed in **cancer** (**Koc**); zipcode binding protein, a protein which binds to a conserved nucleotide element in chicken beta-actin mRNA (ZBP1); and a protein which binds to a promoter cis element in Xenopus laevis TFIIIA gene (B3). p62 protein is cytoplasmic in location, and autoantibodies were found in 21% of a cohort of HCC patients. Patients with chronic hepatitis and liver cirrhosis, conditions which are frequent precursors to HCC, were negative for these autoantibodies, suggesting that the **immune response** might be related to cellular events leading to transformation. However, the possible involvement of p62 autoantigen as a factor in the transformation process remains to be elucidated.

L5 ANSWER 5 OF 8 CAPLUS COPYRIGHT 2003 ACS
IN Wang, Tongtong; Fan, Liqun
TI Compounds and methods for therapy and diagnosis of lung **cancer**
SO PCT Int. Appl., 261 pp.
CODEN: PIXXD2
AB Compds. and methods for the treatment and diagnosis of lung **cancer** are provided. The inventive compds. include polypeptides contg. at least a portion of a lung tumor protein. Vaccines and pharmaceutical compns. for immunotherapy of lung **cancer** comprising such polypeptides, or DNA mols. encoding such polypeptides, are also provided, together with DNA mols. for prepg. the inventive polypeptides.

L5 ANSWER 6 OF 8 CAPLUS COPYRIGHT 2003 ACS
IN Chen, Yao-Tseng; Gure, Ali; Tsang, Solam; Stockert, Elisabeth; Jager, Elke; Knuth, Alexander; Old, Lloyd J.
TI Isolated nucleic acid molecule encoding the **cancer** associated antigens CT7, **KOC**-2 and **KOC**-3, expression of the antigens and application to **cancer** diagnosis and immunotherapy
SO PCT Int. Appl., 45 pp.
CODEN: PIXXD2
AB The invention relates to newly identified **cancer** assocd. antigens, referred to as CT7, **KOC**-2 and **KOC**-3. The invention also relates to observations regarding known mol. **KOC** -1. It has been discovered that each of these mols. provokes antibodies when expressed by a subject. The ramifications of this observation are also a part of this invention.

L5 ANSWER 7 OF 8 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC.
AU Tan, Eng M. (1)
TI Autoantibodies as reporters identifying aberrant cellular mechanisms in tumorigenesis.
SO Journal of Clinical Investigation, (November, 2001) Vol. 108, No. 10, pp. 1411-1415. print.
ISSN: 0021-9738.

L5 ANSWER 8 OF 8 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC.
AU Yamada, Akira (1); Kawano, Koichiro; Harashima, Nanae; Niiya, Fumihiko; Nagai, Kouji; Kobayashi, Terutada; Mine, Takashi; Ushijima, Kimio; Nishida, Takashi; Itoh, Kyogo
TI Study of HLA class I restriction and the directed antigens of cytotoxic T lymphocytes at the tumor sites of ovarian **cancer**.
SO Cancer Immunology Immunotherapy, (May-June, 1999) Vol. 48, No. 2-3, pp. 147-152.
ISSN: 0340-7004.
AB The molecular basis of **T-cell**-mediated recognition of ovarian **cancer** cells remains to be fully addressed. In this

study we investigated HLA class I restriction and directed antigens of cytotoxic T lymphocytes (CTL) at the sites of ovarian **cancer**. Three HLA-class-I-restricted CTL lines were established from the tumor sites of ovarian **cancer** by culturing tumor-infiltrating lymphocytes or tumor-associated ascitic lymphocytes with interleukin-2: (1) HLA-A2402-restricted and ovarian-adenocarcinoma-specific CTL, (2) HLA-A2-restricted CTL recognizing histologically different **cancers**, and (3) HLA-B52-restricted and ovarian-**cancer**-specific CTL. HLA-A0201, HLA-A0206 and HLA-A0207 tumor cells were lysed by the HLA-A2-restricted CTL. HLA-B52 restriction of the third CTL line was confirmed by the transfection of HLA-B5201 cDNA into the tumor cells. The HLA-A2-restricted CTL recognized the SART-1, but not the MAGE-1 or MAGE-3 antigen. These results may facilitate a better understanding of the molecular basis of tumor-specific immunity at the tumor site of ovarian **cancer**.

=> d bib 5 6 l5

L5 ANSWER 5 OF 8 CAPLUS COPYRIGHT 2003 ACS
AN 2000:742121 CAPLUS
DN 133:308983
TI Compounds and methods for therapy and diagnosis of lung **cancer**
IN Wang, Tongtong; Fan, Liqun
PA Corixa Corporation, USA
SO PCT Int. Appl., 261 pp.
CODEN: PIXXD2
DT Patent
LA English
FAN.CNT 12

	PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
PI	WO 2000061612	A2	20001019	WO 2000-US8896	20000403
	WO 2000061612	A3	20010426		
	W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM				
	RW: GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG				
	US 6482597	B1	20021119	US 2000-480884	20000110
	EP 1169347	A2	20020109	EP 2000-920102	20000403
	R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, SI, LT, LV, FI, RO				
	BR 2000009505	A	20020611	BR 2000-9505	20000403
	JP 2002543769	T2	20021224	JP 2000-611554	20000403
PRAI	US 1999-285479	A	19990402		
	US 1999-466396	A	19991217		
	US 1999-476496	A	19991230		
	US 2000-480884	A	20000110		
	US 2000-510376	A	20000222		
	US 1998-40802	A2	19980318		
	US 1998-123912	A2	19980727		
	US 1998-221107	A2	19981222		
	WO 1999-US5798	W	19990317		
	WO 2000-US8896	W	20000403		

L5 ANSWER 6 OF 8 CAPLUS COPYRIGHT 2003 ACS
AN 1999:691302 CAPLUS
DN 131:335828
TI Isolated nucleic acid molecule encoding the **cancer** associated

antigens CT7, KOC-2 and KOC-3, expression of the antigens and application to cancer diagnosis and immunotherapy

IN Chen, Yao-Tseng; Gure, Ali; Tsang, Solam; Stockert, Elisabeth; Jager, Elke; Knuth, Alexander; Old, Lloyd J.

PA Ludwig Institute for Cancer Research, USA

SO PCT Int. Appl., 45 pp.
CODEN: PIXXD2

DT Patent

LA English

FAN.CNT 1

	PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
PI	WO 9954738	A1	19991028	WO 1999-US5766	19990316
	W: AU, CA, CN, JP, KR, NZ, ZA				
	RW: AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE				
	US 6297364	B1	20011002	US 1998-61709	19980417
	CA 2324975	AA	19991028	CA 1999-2324975	19990316
	AU 9930079	A1	19991108	AU 1999-30079	19990316
	AU 745259	B2	20020314		
	EP 1071957	A1	20010131	EP 1999-911439	19990316
	R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, FI				
	JP 2002512049	T2	20020423	JP 2000-545030	19990316
	US 2002111470	A1	20020815	US 2001-899651	20010706
PRAI	US 1998-61709	A	19980417		
	WO 1999-US5766	W	19990316		

RE.CNT 5 THERE ARE 5 CITED REFERENCES AVAILABLE FOR THIS RECORD
ALL CITATIONS AVAILABLE IN THE RE FORMAT

=>

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: April 24, 2003, 16:11:07 ; Search time 38 Seconds
(without alignments)
3139.507 Million cell updates/sec

Title: US-09-897-778-176
Perfect score: 2956
Sequence: 1 MNKLYIGNLSENAAPSDLES.....VKQHQQQKALQSGPPQSRRK 579

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2956	100.0	579	4	O00425	O00425 homo sapien
2	2858	96.7	579	11	Q9CPN8	Q9cpn8 mus musculu
3	2483	84.0	593	13	O57526	O57526 xenopus lae
4	2461.5	83.3	594	13	O73932	O73932 xenopus lae
5	2261.5	76.5	582	13	Q9PW80	Q9pw80 brachydanio
6	2219.5	75.1	576	13	O42254	O42254 gallus gall
7	2192	74.2	577	11	O88477	O88477 mus musculu
8	2181	73.8	577	4	Q9NZI8	Q9nzi8 homo sapien
9	1868	63.2	556	4	Q9Y6M1	Q9y6m1 homo sapien
10	835.5	28.3	558	5	Q9VZ69	Q9vz69 drosophila
11	835	28.2	566	5	Q9NGS9	Q9ngs9 drosophila
12	473	16.0	100	11	Q9D054	Q9d054 mus musculu
13	418.5	14.2	854	5	Q21605	Q21605 caenorhabdi
14	262.5	8.9	774	10	Q9LI28	Q9li28 oryza sativ
15	260.5	8.8	542	10	Q8S7G1	Q8s7g1 oryza sativ
16	242.5	8.2	557	5	Q23487	Q23487 caenorhabdi
17	240	8.1	644	4	Q12828	Q12828 homo sapien
18	240	8.1	651	11	Q91WJ8	Q91wj8 mus musculu
19	238	8.1	641	5	Q9BLA0	Q9bla0 caenorhabdi
20	236.5	8.0	568	10	Q9LXF5	Q9lxf5 arabidopsis
21	235.5	8.0	621	10	Q9C553	Q9c553 arabidopsis
22	235.5	8.0	653	4	Q96AE4	Q96ae4 homo sapien
23	234	7.9	600	4	Q92946	Q92946 homo sapien
24	233	7.9	492	4	Q9UNW9	Q9unw9 homo sapien
25	233	7.9	498	4	O43267	O43267 homo sapien
26	231.5	7.8	313	10	Q9LZ82	Q9lz82 arabidopsis
27	231	7.8	364	5	Q95SZ9	Q95sz9 drosophila
28	230.5	7.8	493	11	Q9JKN6	Q9jkn6 mus musculu
29	230	7.8	618	5	Q9BLA1	Q9bla1 caenorhabdi
30	229.5	7.8	398	3	O74919	O74919 schizosacch
31	229	7.7	386	5	P91632	P91632 drosophila
32	228.5	7.7	313	4	Q96EP6	Q96ep6 homo sapien
33	226	7.6	1301	5	Q9U982	Q9u982 drosophila
34	226	7.6	1301	5	Q9V8H6	Q9v8h6 drosophila
35	225.5	7.6	577	10	Q9SR13	Q9sr13 arabidopsis
36	225.5	7.6	589	5	Q17935	Q17935 caenorhabdi
37	225.5	7.6	611	5	Q17936	Q17936 caenorhabdi
38	224	7.6	769	13	Q8UVD9	Q8uvd9 gallus gall
39	222	7.5	479	10	Q9XI71	Q9xi71 arabidopsis
40	219.5	7.4	396	13	Q9PSI1	Q9psi1 xenopus lae
41	217.5	7.4	640	10	Q9ASX3	Q9asx3 arabidopsis
42	217.5	7.4	644	10	Q9FNK3	Q9fnk3 arabidopsis
43	215.5	7.3	403	11	Q9DB01	Q9db01 mus musculu
44	214	7.2	833	10	Q9FMF0	Q9fmf0 arabidopsis
45	213.5	7.2	510	10	Q8W5C2	Q8w5c2 oryza sativ

ALIGNMENTS

RESULT 1
O00425
ID O00425 PRELIMINARY; PRT; 579 AA.
AC O00425;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative RNA binding protein KOC (KOC).
GN KOC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Mueller-Pillasch F., Lacher U., Wallrapp C., Et AL.;
RL Oncogene 0:0-0(0).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=PANCREAS;
RA Mueller-Pillasch F., Lacher U., Wallrapp C.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
DR EMBL; U97188; AAC35208.1; -.
DR EMBL; U76705; AAD09223.1; -.
DR InterPro; IPR004087; KH_dom.
DR InterPro; IPR004088; KH_type_1.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00013; KH-domain; 4.
DR Pfam; PF00076; rrm; 2.
DR SMART; SM00322; KH; 4.
DR SMART; SM00360; RRM; 2.
DR PROSITE; PS50084; KH_TYPE_1; 4.
DR PROSITE; PS50102; RRM; 2.
DR PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
SQ SEQUENCE 579 AA; 63720 MW; AE5C3A8EE3C135C5 CRC64;

Query Match 100.0%; Score 2956; DB 4; Length 579;
Best Local Similarity 100.0%; Pred. No. 1.7e-187;
Matches 579; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGK 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGK 60

Qy      61 IELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 IELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETA 120

Qy     121 VVNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDEMAAQQNPLQQPRGRRGLGQRGSS 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 VVNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDEMAAQQNPLQQPRGRRGLGQRGSS 180

Qy     181 RQGSPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAA 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 RQGSPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAA 240

Qy     241 EKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEGRNLK 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 EKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEGRNLK 300

Qy     301 KIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIASMNL 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 KIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIASMNL 360

Qy     361 QAHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMTPPYPQFEQSETETVHQFIPALSVGAI 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 QAHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMTPPYPQFEQSETETVHQFIPALSVGAI 420

Qy     421 IGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEENFV 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 IGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEENFV 480

Qy     481 SPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVVVKIT 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 SPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVVVKIT 540

Qy     541 GHFYACQVAQRKIQEILTQVKQHQQQKALQSGPPQSRRK 579
           |||||||||||||||||||||||||||||||||||||||
Db     541 GHFYACQVAQRKIQEILTQVKQHQQQKALQSGPPQSRRK 579
```

```
RESULT 2
Q9CPN8
ID   Q9CPN8      PRELIMINARY;      PRT;   579 AA.
AC   Q9CPN8;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   10 days embryo cDNA, RIKEN full-length enriched library,
DE   clone:2610036B18, full insert sequence (Igf2 mRNA-binding protein
DE   3).
GN   IGF2BP3 OR 2610101N11RIK OR MIMP3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=EMBRYO;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Justincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Mori H., Sakakibara S., Imai T., Nakamura Y., Iijima T., Suzuki A.,
RA   Yuasa Y., Takeda M., Okano H.;
RT   "Expression of mouse igf2 mRNA-binding protein 3 and its implications
RT   for the developing central nervous system.";
RL   J. Neurosci. Res. 0:0-0(2001).
DR   EMBL; AK011689; BAB27779.1; -.
DR   EMBL; AB046173; BAB19755.1; -.
DR   MGD; MGI:1890359; Igf2bp3.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00013; KH-domain; 4.
DR   Pfam; PF00076; rrm; 2.
DR   SMART; SM00322; KH; 4.
DR   SMART; SM00360; RRM; 2.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
DR   PROSITE; PS50102; RRM; 2.
SQ   SEQUENCE   579 AA;  63574 MW;  CABD9A4355B392B7 CRC64;

  Query Match             96.7%;  Score 2858;  DB 11;  Length 579;
  Best Local Similarity   96.4%;  Pred. No. 5.2e-181;
  Matches  558;  Conservative   10;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGK 60
           ||||||||||::| |:||||:||||||||:||||||||||||||||| ||||||||||||
Db       1 MNKLYIGNLSDHAGPADLESVFKDAKIPVAGPFLVKTGYAFVDCPDEGWALKAIEALSGK 60

Qy      61 IELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETA 120
           :||||||:||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 MELHGKPMEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETA 120

Qy     121 VVNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDEMAAQQNPLQQPRGRRGLGQRGSS 180
           |||||||||||||||||||||||||||||||||||| |||||  | ||||| ||||||
Db     121 VVNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDETAAQQNPSPQLRGRRGPGQRGSS 180

Qy     181 RQGSPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAA 240
           || ||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db     181 RQASPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENTGAA 240

Qy     241 EKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEGRNLK 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 EKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEGRNLK 300

Qy     301 KIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIASMNL 360
           ||||||||||||||||||||||||||||||||||||:||||||||||||||||||||||||
Db     301 KIEQDTDTKITISPLQELTLYNPERTITVKGSVETCAKAEEEIMKKIRESYENDIASMNL 360

Qy     361 QAHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMTPPYPQFEQSETETVHQFIPALSVGAI 420
           |||||||||||||||||||||||||||||| :|||||||||||||| ||||||||||
Db     361 QAHLIPGLNLNALGLFPPTSGMPPPTSGPPSTLTPPYPQFEQSETETVHLFIPALSVGAI 420

Qy     421 IGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEENFV 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 IGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEENFV 480

Qy     481 SPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVVVKIT 540
           ||||||||||||||||||||||||||||||||||||:||||||||||||||||||||||||
Db     481 SPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQSLSSAEVVVPRDQTPDENDQVVVKIT 540

Qy     541 GHFYACQVAQRKIQEILTQVKQHQQQKALQSGPPQSRRK 579
           |||||||||||||||||||||||||||||||||||||||
Db     541 GHFYACQVAQRKIQEILTQVKQHQQQKALQSGPPQSRRK 579
```

```
RESULT 3
O57526
ID   O57526      PRELIMINARY;      PRT;   593 AA.
AC   O57526;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   KH domain-containing transcription factor B3.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92249652; PubMed=1577195;
RA   Pfaff S.L., Taylor W.L.;
RT   "Characterization of a Xenopus oocyte factor that binds to a
RT   developmentally regulated cis-element in the TFIIIA gene.";
RL   Dev. Biol. 151:306-316(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Griffin D., Taylor W.L.;
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Havin L., Git A., Elisha Z., Oberman F., Yaniv K.,
RA   Pressman Schwartz S., Standart N.M., Yisraeli J.K.;
RL   Genes Dev. 0:0-0(1998).
DR   EMBL; AF042353; AAB97457.1; -.
DR   EMBL; AF064633; AAC18597.1; -.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00013; KH-domain; 4.
DR   Pfam; PF00076; rrm; 2.
DR   SMART; SM00322; KH; 4.
DR   SMART; SM00360; RRM; 2.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
DR   PROSITE; PS50102; RRM; 2.
DR   PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
SQ   SEQUENCE   593 AA;  65385 MW;  5A5AB4B4A1D55DF7 CRC64;

  Query Match           84.0%;  Score 2483;  DB 13;  Length 593;
  Best Local Similarity 82.1%;  Pred. No. 3.6e-156;
  Matches  487;  Conservative   47;  Mismatches   45;  Indels   14;  Gaps    6;

Qy       1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGK 60
           |||||||||||| :| |||||:|||::||| :| ||||:|||||||||:||:|||: ||||
Db       1 MNKLYIGNLSENVSPPDLESLFKESKIPFTGQFLVKSGYAFVDCPDETWAMKAIDTLSGK 60

Qy      61 IELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETA 120
           :||||| |||||||||||| ||||||||||||||||||||| ||| ||:|||||||||||
Db      61 VELHGKVIEVEHSVPKRQRSRKLQIRNIPPHLQWEVLDSLLAQYGTVENCEQVNTDSETA 120

Qy     121 VVNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDEMAAQQNPLQQ--------PRGRR 172
           |||||||:|: ||| |:||||:||||::||| ||||||| |:| ||         |:|||
Db     121 VVNVTYANKEHARQGLEKLNGYQLENYSLKVTYIPDEMATPQSPSQQLQQPQQQHPQGRR 180

Qy     173 GLGQRGSSRQGSPGSVSKQKP-CDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDV 231
           | |||| :|||||: :: ||  ::|||:|||||||||||||||||||||||||||||||:
Db     181 GFGQRGPARQGSPGAAARPKPQSEVPLRMLVPTQFVGAIIGKEGATIRNITKQTQSKIDI 240

Qy     232 HRKENAGAAEKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRL 291
           |||||||||||| ||| |||:| |||:| |:||| |||:| |||||||||||||||||||
Db     241 HRKENAGAAEKPITIHSTPEGCSAACKIIMEIMQKEAQDTKFTEEIPLKILAHNNFVGRL 300

Qy     292 IGKEGRNLKKIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESY 351
           |||||||||||||||||||||||||:|||||||||||||::|||||||||:||||||||
Db     301 IGKEGRNLKKIEQDTDTKITISPLQDLTLYNPERTITVKGSIETCAKAEEEVMKKIRESY 360

Qy     352 ENDIASMNLQAHLIPGLNLNALGLFPP-TSGMPPPTSGPPSAMT-PPYPQF-EQSETETV 408
           |||||:|||||||||||||||||||||| :|||||::|  |  |   || | :| |:|||
Db     361 ENDIAAMNLQAHLIPGLNLNALGLFPPSSSGMPPPSAGVSSPTTSASYPPFGQQPESETV 420

Qy     409 HQFIPALSVGAIIGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQG 468
           | ||||:|||||||||||||||||||||||||||||| ||||:|||||||||||||||
Db     421 HLFIPALAVGAIIGKQGQHIKQLSRFAGASIKIAPAEGPDAKLRMVIITGPPEAQFKAQG 480

Qy     469 RIYGKIKEENFVSPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQT 528
           |||||:|||||  ||||||||||:|||:|||||||||||||||||:||||||||||
Db     481 RIYGKLKEENFFGPKEEVKLEAHIKVPSYAAGRVIGKGGKTVNELQNLTSAEVVVPRDQT 540

Qy     529 PDENDQVVVKITGHFYACQVAQRKIQEILTQVK--QHQQQKALQSGPPQSRRK 579
           |||||||||||||||| |:|||||||||| ||:  | ||||  ||| || |||
Db     541 PDENDQVVVKITGHFYASQLAQRKIQEILAQVRRQQQQQQKTAQSGQPQPRRK 593


RESULT 4
O73932
ID   O73932      PRELIMINARY;      PRT;   594 AA.
AC   O73932;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   VG1 RNA binding protein variant D.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Havin L., Git A., Elisha Z., Oberman F., Yaniv K.,
RA   Pressman Schwartz S., Standart N.M., Yisraeli J.K.;
RL   Genes Dev. 0:0-0(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98228351; PubMed=9560341;
RA   Deshler J.O., Highett M.I., Abramson T., Schnapp B.J.;
RT   "A highly conserved RNA-binding protein for cytoplasmic mRNA
RT   localization in vertebrates.";
RL   Curr. Biol. 8:489-496(1998).
DR   EMBL; AF064634; AAC18598.1; -.
DR   EMBL; AF055923; AAC41285.1; -.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00013; KH-domain; 4.
DR   Pfam; PF00076; rrm; 2.
DR   SMART; SM00322; KH; 4.
DR   SMART; SM00360; RRM; 2.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
DR   PROSITE; PS50102; RRM; 2.
SQ   SEQUENCE   594 AA;  65643 MW;  54CEA7BFF0856DD6 CRC64;

  Query Match           83.3%;  Score 2461.5;  DB 13;  Length 594;
  Best Local Similarity 81.3%;  Pred. No. 9.4e-155;
  Matches  483;  Conservative   49;  Mismatches   47;  Indels   15;  Gaps    6;

Qy       1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGK 60
           |||||||||||| :|:||||:|::||| :| ||||:|||||||||:||:|||: ||||
Db       1 MNKLYIGNLSENVSPTDLESLFKESKIPFTGQFLVKSGYAFVDCPDETWAMKAIDTLSGK 60

Qy      61 IELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETA 120
           :||||| |||||||||||| ||||||||||||||||||||| ||| ||:||||||:||||
Db      61 VELHGKVIEVEHSVPKRQRSRKLQIRNIPPHLQWEVLDSLLAQYGTVENCEQVNTESETA 120

Qy     121 VVNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDEMAAQQNPLQQ---------PRGR 171
           |||||||::|: ||| |:||||:||||::||| |||||||  | | ||         |:||
Db     121 VVNVTYANKEHARQGLEKLNGYQLENYSLKVTYIPDEMATPQAPSQQLQQQPQQQHPQGR 180

Qy     172 RGLGQRGSSRQGSPGSVSKQKP-CDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKID 230
           || |||| :|||||: :: ||  ::|||:|||||||||||||||||||||||||||||||
Db     181 RGFGQRGPARQGSPGAAARPKPQTEVPLRMLVPTQFVGAIIGKEGATIRNITKQTQSKID 240
```

```
Qy    231 VHRKENAGAAEKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGR 290
          :||||||||||| ||| ||||| ||||| |:||| ||||| ||||||||||||||||||||
Db    241 IHRKENAGAAEKPITIHSTPEGCSAACKIIMEIMQKEAQDTKFTEEIPLKILAHNNFVGR 300

Qy    291 LIGKEGRNLKKIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRES 350
          |||||||||||||||||||||||||:|||||||||||||||::| ||||||||||||||||
Db    301 LIGKEGRNLKKIEQDTDTKITISPLQDLTLYNPERTITVKGSIEPCAKAEEEIMKKIRES 360

Qy    351 YENDIASMNLQAHLIPGLNLNALGLFP-PTSGMPPPTSGPPS-AMTPPYPQF-EQSETET 407
          ||||||:|||||||||||||||||||  :|||||||: | ||   :  || | :| |:||
Db    361 YENDIAAMNLQAHLIPGLNLNALGLFPSSSSGMPPPSVGVPSPTSSTSYPPFGQQPESET 420

Qy    408 VHQFIPALSVGAIIGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQ 467
          || |||||:|||||||||||||||||||||||||||||| ||||:|||||||||||||||
Db    421 VHLFIPALAVGAIIGKQGQHIKQLSRFAGASIKIAPAEGPDAKLRMVIITGPPEAQFKAQ 480

Qy    468 GRIYGKIKEENFVSPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQ 527
          ||||||:|||||  ||||||||| ||:|||:|||||||||||||||||||:||||||||||
Db    481 GRIYGKLKEENFFGPKEEVKLETHIKVPSYAAGRVIGKGGKTVNELQNLTSAEVVVPRDQ 540

Qy    528 TPDENDQVVVKITGHFYACQVAQRKIQEILTQVK--QHQQQKALQSGPPQSRRK 579
          ||||||:|||||||||||| |:|||||||||| ||:  | |||| :||| || |||
Db    541 TPDENDEVVVKITGHFYASQLAQRKIQEILAQVRRQQQQQQKTVQSGQPQPRRK 594


RESULT 5
Q9PW80
ID   Q9PW80      PRELIMINARY;      PRT;   582 AA.
AC   Q9PW80;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Vg1 RNA binding protein.
GN   DVR1RBP.
OS   Brachydanio rerio (Zebrafish) (Zebra danio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Zhang Q., Yaniv K., Oberman F., Wolke U., Git A., Fromer M.,
RA   Taylor W., Meyer D., Standart N., Raz E., Yisraeli J.K.;
RT   "Vg1 RBP intracellular distribution and evolutionarily conserved
RT   expression suggest multiple roles during development.";
RL   Mech. Dev. 0:0-0(1999).
DR   EMBL; AF161270; AAD45610.1; -.
DR   ZFIN; ZDB-GENE-000308-1; dvr1rbp.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00013; KH-domain; 4.
DR   Pfam; PF00076; rrm; 2.
DR   SMART; SM00322; KH; 4.
DR   SMART; SM00360; RRM; 2.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
DR   PROSITE; PS50102; RRM; 2.
SQ   SEQUENCE   582 AA;  63351 MW;  9DAE63200681B306 CRC64;

  Query Match           76.5%;  Score 2261.5;  DB 13;  Length 582;
  Best Local Similarity   77.1%;  Pred. No. 1.6e-141;
  Matches  450;  Conservative   50;  Mismatches   77;  Indels    7;  Gaps    5;

Qy      1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGK 60
          |||||||||:|| |:  ||||||:  ||| | |||||:|||||||||  |::||: ||||
Db      1 MNKLYIGNVSEQASALDLESIFEQWKIPFSAPFLVKSGYAFVDCPDEKVAMRAIDTLSGK 60

Qy     61 IELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETA 120
          :||||| :||||||||||  ||||||||||:|||||| || ||| ||||||||||:|||
Db     61 VELHGKVLEVEHSVPKRQRSCKLQIRNIPPHMQWEVLDGLLAQYGTVESCEQVNTDTETA 120

Qy    121 VVNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDEMAAQQNPLQQPRGRRGLGQRGSS 180
```

```
          |||| | :|||||:|:|||||| :||: |||:||||| ||   |     ||||   ||
Db    121 VVNVRYGAKDQAREAMDKLNGFLMENYALKVSYIPDETAAADAP--AVGGRRGFNPRGPP 178

Qy    181 RQGSPGSVSKQK-PCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGA 239
          |||||   :: |   |:|||||||||||||||||||||||||||| ||||:||||||||
Db    179 RQGSPSLGARPKLQSDVPLRLLVPTQFVGAIIGKEGATIRNITKQTHSKIDIHRKENAGA 238

Qy    240 AEKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEGRNL 299
          ||| ||: ||||| |:||::|:||| ||| | | ||||||||||||||||||||||||||
Db    239 AEKPITVHSTPEGCSSACRNIMEIMQKEAIDTKITEEIPLKILAHNNFVGRLIGKEGRNL 298

Qy    300 KKIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIASMN 359
          ||||||||||||||||:|||||||||||||| :: ||||||||||:|||||||:|:|:
Db    299 KKIEQDTDTKITISPLQDLTLYNPERTITVKGTLDACAKAEEEIMKKVRESYENDVAAMH 358

Qy    360 LQAHLIPGLNLNALGLFP-PTSG--MPPPTSGPPSAMTPPYPQF-EQSETETVHQFIPAL 415
          ||::|||||||||||||||  ||   |  ||||     |  |  | |:|||| |||||
Db    359 LQSNLIPGLNLNALGLFPGAASGGISPSVVSGPPPGAQAGYQSFGAQMESETVHLFIPAL 418

Qy    416 SVGAIIGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIK 475
          :|||||||||||||||||||||||||||||:  ||| |||||:|||||||||||:||:|
Db    419 AVGAIIGKQGQHIKQLSRFAGASIKIAPADGIDAKQRMVIISGPPEAQFKAQGRIFGKLK 478

Qy    476 EENFVSPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQV 535
          ||||  |||||||||||:|||||||||||||||||||||:|||||||||||||||||||
Db    479 EENFFGPKEEVKLEAHIKVPSFAAGRVIGKGGKTVNELQNLTSAEVVVPRDQTPDENDQV 538

Qy    536 VVKITGHFYACQVAQRKIQEILTQVKQHQQQKALQSGPPQSRRK 579
          |||||||||| |:|||||||::|||:: || |   :||| :|||
Db    539 VVKITGHFYASQLAQRKIQEIISQVRRQQQPKPSAAGPPVARRK 582


RESULT 6
O42254
ID   O42254      PRELIMINARY;      PRT;   576 AA.
AC   O42254;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Zipcode-binding protein.
GN   ZBP1.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97220007; PubMed=9121465;
RA   Ross A.F., Oleynikov Y.S., Kislauskis E.H., Taneja K.L., Singer R.H.;
RT   "Characterization of a beta-actin mRNA zipcode-binding protein.";
RL   Mol. Cell. Biol. 17:2158-2165(1997).
DR   EMBL; AF026527; AAB82295.1; -.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00013; KH-domain; 4.
DR   Pfam; PF00076; rrm; 2.
DR   SMART; SM00322; KH; 4.
DR   SMART; SM00360; RRM; 2.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
DR   PROSITE; PS50102; RRM; 2.
DR   PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
SQ   SEQUENCE   576 AA;  63271 MW;  01AAF2D1D81C8811 CRC64;

  Query Match           75.1%;  Score 2219.5;  DB 13;  Length 576;
  Best Local Similarity   74.7%;  Pred. No. 9.4e-139;
  Matches  437;  Conservative   61;  Mismatches   72;  Indels   15;  Gaps    7;

Qy      1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGK 60
          ||||||||||:|:  |:||| :| | ||  || ||||:|||||||||| ||:|||  |||
Db      1 MNKLYIGNLNESVTPADLEKVFNDHKISFSGQFLVKSGYAFVDCPDEQWAMKAIETFSGK 60
```

```
Qy      61 IELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETA 120
           :|||||  :|:||||||:|| ||:|||||| |:|||||  || ||| ||:|||||||||||
Db      61 VELHGKQLEIEHSVPKKQRSRKIQIRNIPPQLRWEVLDGLLAQYGTVENCEQVNTDSETA 120

Qy     121 VVNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDEMAAQQNPLQQPRGRR-GLGQRGS 179
           ||||||:::|  |||: |||| |||| |||:||||| : |        ||| | | ||:
Db     121 VVNVTYTNREQTRQAIMKLNGHQLENHVLKVSYIPDEQSVQ----GPENGRRGGFGARGA 176

Qy     180 SRQGSP---GSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKEN 236
           |||||   |:  ||:| |:||||||||:|||||||||||||||||||||||||||||||
Db     177 PRQGSPVTAGAPVKQQPVDIPLRLLVPTQYVGAIIGKEGATIRNITKQTQSKIDVHRKEN 236

Qy     237 AGAAEKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEG 296
           ||||||:|:| ||||| ||||| ||||| |||:| |  :|:||||||||||||||||||
Db     237 AGAAEKAISIHSTPEGCSAACKMILEIMQKEAKDTKTADEVPLKILAHNNFVGRLIGKEG 296

Qy     297 RNLKKIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIA 356
           |||||:||||:|||||| ||:|||||||||||||||::| | |||:|||||:||:||||:|
Db     297 RNLKKVEQDTETKITISSLQDLTLYNPERTITVKGSIENCCKAEQEIMKKVREAYENDVA 356

Qy     357 SMNLQAHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMT--PPYPQFEQSETETVHQFIPA 414
           :|:||:|||||||| |:|||| :|   ||   |||::: : ||  |   | |||| ||||
Db     357 AMSLQSHLIPGLNLAAVGLFPASSNAVPP---PPSSVSGAAPYSSFMPPEQETVHVFIPA 413

Qy     415 LSVGAIIGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKI 474
           :|||||||:|||||||||| |||||||  | ||:|||||:|||||||||||||||||:|:
Db     414 QAVGAIIGKKGQHIKQLSRFASASIKIAPPETPDSKVRMVVITGPPEAQFKAQGRIYGKL 473

Qy     475 KEENFVSPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQ 534
           |||||  |||||||| |||||: |||||||||||||||||||::|||||||||||||:|
Db     474 KEENFFGPKEEVKLETHIRVPASAAGRVIGKGGKTVNELQNLTAAEVVVPRDQTPDENEQ 533

Qy     535 VVVKITGHFYACQVAQRKIQEILTQVKQHQQQKALQSGPPQSRRK 579
           |:||| ||||| |:|||||::|| |||| | ||  |||  |:|||
Db     534 VIVKIIGHFYASQMAQRKIRDILAQVKQ-QHQKG-QSGQLQARRK 576
```

RESULT 7
O88477
ID O88477 PRELIMINARY; PRT; 577 AA.
AC O88477;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Coding region determinant binding protein (Coding region determinant-
DE binding protein).
GN IGF2BP1 OR CRDBP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92217743; PubMed=1559612;
RA Bernstein P.L., Herrick D.J., Prokipcak R.D., Ross J.;
RT "Control of c-myc mRNA half-life in vitro by a protein capable of
RT binding to a coding region stability determinant.";
RL Genes Dev. 6:642-654(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94158886; PubMed=8114742;
RA Herrick D.J., Ross J.;
RT "The half-life of c-myc mRNA in growing and serum-stimulated cells:
RT influence of the coding and 3' untranslated regions and role of
RT ribosome translocation.";
RL Mol. Cell. Biol. 14:2119-2128(1994).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=94179348; PubMed=8132663;
RA Prokipcak R.D., Herrick D.J., Ross J.;
RT "Purification and properties of a protein that binds to the C-terminal
RT coding region of human c-myc mRNA.";
RL J. Biol. Chem. 269:9261-9269(1994).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=97322234; PubMed=9178888;
RA Leeds P., Kren B.T., Boylan J.M., Betz N.A., Steer C.J.,
RA Gruppuso P.A., Ross J.;
RT "Developmental regulation of CRD-BP, an RNA-binding protein that
RT stabilizes c-myc mRNA in vitro.";
RL Oncogene 14:1279-1286(1997).
RN [5]
RP SEQUENCE FROM N.A.
RA Ross J., Prokipcak R.D., Leeds P., Doyle G.A.R., Betz N.A.,
RA Fleisig A.J.;
RL Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYONIC HEAD;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AF061569; AAC72743.1; -.
DR EMBL; AK013940; BAB29071.1; -.
DR HSSP; P11940; 1CVJ.
DR MGD; MGI:1890357; Igf2bp1.
DR InterPro; IPR004087; KH_dom.
DR InterPro; IPR004088; KH_type_1.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00013; KH-domain; 4.
DR Pfam; PF00076; rrm; 2.
DR SMART; SM00322; KH; 4.
DR SMART; SM00360; RRM; 2.
DR PROSITE; PS50084; KH_TYPE_1; 4.
DR PROSITE; PS50102; RRM; 2.
SQ SEQUENCE 577 AA; 63450 MW; EFBB1AF2FF9F0344 CRC64;

Query Match 74.2%; Score 2192; DB 11; Length 577;
Best Local Similarity 74.1%; Pred. No. 6.2e-137;
Matches 434; Conservative 62; Mismatches 74; Indels 16; Gaps 8;

```
Qy       1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGK 60
           |||||||||:|:  |:||| :| : ||  || ||||:|||||||||| ||:|||| |||
Db       1 MNKLYIGNLNESVTPADLEKVFAEHKISYSGQFLVKSGYAFVDCPDEHWAMKAIETFSGK 60

Qy      61 IELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETA 120
           :|| || :|:|||||||:|| ||:||||||| |:|||||||| ||| ||:|||||:||||
Db      61 VELQGKRLEIEHSVPKKQRSRKIQIRNIPPQLRWEVLDSLLAQYGTVENCEQVNTESETA 120

Qy     121 VVNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDEMAAQQNPLQQPRGRR-GLGQRGS 179
           ||||||||:::| |||: |||| |||| |||:|||||   |        ||| | | ||
Db     121 VVNVTYSNREQTRQAIMKLNGHQLENHALKVSYIPDEQITQ----GPENGRRGGFGSRGQ 176

Qy     180 SRQGSP---GSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKEN 236
           |||||   |: :||:| |:|||||||||:|||||||||||||||||||||||||||||||
Db     177 PRQGSPVAAGAPAKQQPVDIPLRLLVPTQYVGAIIGKEGATIRNITKQTQSKIDVHRKEN 236

Qy     237 AGAAEKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEG 296
```

```
Db    237 AGAAEKAISVHSTPEGCSSACKMILEIMHKEAKDTKTADEVPLKILAHNNFVGRLIGKEG 296

Qy    297 RNLKKIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIA 356

Db    297 RNLKKVEQDTETKITISSLQDLTLYNPERTITVKGAIENCCRAEQEIMKKVREAYENDVA 356

Qy    357 SMNLQAHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMT--PPYPQFEQS-ETETVHQFIP 413

Db    357 AMSLQSHLIPGLNLAAVGLFPASSSAVPP---PPSSVTGAAPYSSFMQAPEQEMVQVFIP 413

Qy    414 ALSVGAIIGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGK 473

Db    414 AQAVGAIIGKKGQHIKQLSRFASASIKIAPPETPDSKVRMVVITGPPEAQFKAQGRIYGK 473

Qy    474 IKEENFVSPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDEND 533

Db    474 LKEENFFGPKEEVKLETHIRVPASAAGRVIGKGGKTVNELQNLTAAEVVVPRDQTPDEND 533

Qy    534 QVVVKITGHFYACQVAQRKIQEILTQVKQHQQQKALQSGPPQSRRK 579

Db    534 QVIVKIIGHFYASQMAQRKIRDILAQVKQ-QHQKG-QSNLAQARRK 577


RESULT 8
Q9NZI8
ID   Q9NZI8      PRELIMINARY;      PRT;   577 AA.
AC   Q9NZI8;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   mRNA-binding protein CRDBP.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ioannidis P., Trangas T., Dimitriadis E., Samiotaki M.,
RA   Panoutsakopoulos G., Kyriazoglou I., Voutzoulias S., Tsiapalis C.M.,
RA   Kittas C., Agnantis N., Pandis N.;
RT   "Ectopic expression of a KH-domain containing protein, highly
RT   homologous to both human IMP-1 and mouse CRD-BP, in benign and
RT   malignant mesenchymal tumors.";
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF198254; AAF37203.1; -.
DR   HSSP; P11940; 1CVJ.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00013; KH-domain; 4.
DR   Pfam; PF00076; rrm; 2.
DR   SMART; SM00322; KH; 4.
DR   SMART; SM00360; RRM; 2.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
DR   PROSITE; PS50102; RRM; 2.
SQ   SEQUENCE   577 AA;  63456 MW;  0749A060F252D81D CRC64;

  Query Match           73.8%;  Score 2181;  DB 4;  Length 577;
  Best Local Similarity   73.9%;  Pred. No. 3.3e-136;
  Matches  433;  Conservative   61;  Mismatches   76;  Indels   16;  Gaps    8;

Qy      1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGK 60

Db      1 MNKLYIGNLNESVTPADLEKVFAEHKISYSGQFLVKSGYAFVDCPDEHWAMKAIETFSGK 60

Qy     61 IELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETA 120

Db     61 VELQGKRLEIEHSVPKKQRSRKIQIRNIPPQLRWEVLDSLLAQYGTVENCEQVNTESETA 120

Qy    121 VVNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDEMAAQQNPLQQPRGRR-GLGQRGS 179

Db    121 VVNVTYSNREQTRQAIMKLNGHQLENHALKVSYIPDEQIAQ----GPENGRRGGFGSRGQ 176

Qy    180 SRQGSP---GSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKEN 236

Db    177 PRQGSPVAAGAPAKQQQVDIPLRLLVPTQYVGAIIGKEGATIRNITKQTQSKIDVHRKEN 236

Qy    237 AGAAEKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEG 296

Db    237 AGAAEKAISVHSTPEGCSSACKMILEIMHKEAKDTKTADEVPLKTLAHNNFVGRLIGKEG 296

Qy    297 RNLKKIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIA 356

Db    297 RNLKKVEQDTETKITISSLQDLTLYNPERTITVKGAIENCCRAEQEIMKKVREAYENDVA 356

Qy    357 SMNLQAHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMT--PPYPQFEQS-ETETVHQFIP 413

Db    357 AMSLQSHLTPGLNLAAVGLFPASSSAVPP---PPSSVTGAAPYSSFMQAPEQEMVQVFIP 413

Qy    414 ALSVGAIIGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGK 473

Db    414 AQAVGAIIGKKGQHIKQLSRFASASIKIAPPETPDSKVRMVIITGPPEAQFKAQGRIYGK 473

Qy    474 IKEENFVSPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDEND 533

Db    474 LKEENFFGPKEEVKLETHIRVPASAAGRVIGKGGKTVNELQNLTAAEVVVPRDQTPDEND 533

Qy    534 QVVVKITGHFYACQVAQRKIQEILTQVKQHQQQKALQSGPPQSRRK 579

Db    534 QVIVKIIGHFYASQMAQRKIRDILAQVKQ-QHQKG-QSNQAQARRK 577


RESULT 9
Q9Y6M1
ID   Q9Y6M1      PRELIMINARY;      PRT;   556 AA.
AC   Q9Y6M1;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hepatocellular carcinoma autoantigen.
GN   P62.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99207072; PubMed=10190901;
RA   Zhang J.Y., Chan E.K., Peng X.X., Tan E.M.;
RT   "A novel cytoplasmic protein with RNA-binding motifs is an autoantigen
RT   in human hepatocellular carcinoma.";
RL   J. Exp. Med. 189:1101-1110(1999).
DR   EMBL; AF057352; AAD31596.1; -.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00013; KH-domain; 4.
DR   Pfam; PF00076; rrm; 2.
DR   SMART; SM00322; KH; 4.
DR   SMART; SM00360; RRM; 2.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
DR   PROSITE; PS50102; RRM; 2.
SQ   SEQUENCE   556 AA;  61842 MW;  1EDEFB100443DDC4 CRC64;

  Query Match           63.2%;  Score 1868;  DB 4;  Length 556;
  Best Local Similarity   64.8%;  Pred. No. 1.6e-115;
  Matches  368;  Conservative   78;  Mismatches   96;  Indels   26;  Gaps    6;

Qy      1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGK 60

Db      2 MNKLYIGNLSPAVTADDLRQLFGDRKLPLAGQVLLKSGYAFVDYPDQNWAIRAIETLSGK 61

Qy     61 IELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETA 120

Db     62 VELHGKIMEVDYSVSKKLRSRKIQIRNIPPHLQWEVLDGLLAQYGTVENVEQVNTDTETA 121
```

```
Qy     121 VVNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDEMAAQQNPLQQPRGRRGLGQRGSS 180
           ||||||:::::|: |::||:| | ||:: |::||||| :  :| |  | :|  |   |
Db     122 VVNVTYATREEAKIAMEKLSGHQFENYSFKISYIPDEEVSSPSPPQ--RAQR--GDHSSR 177

Qy     181 RQG-SPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGA 239
           || :|| |: :  | |||:|||||||||||||||| ||:||||||||::|:|||||:||
Db     178 EQGHAPGGTSQARQIDFPLRILVPTQFVGAIIGKEGLTIKNITKQTQSRVDIHRKENSGA 237

Qy     240 AEKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEGRNL 299
           ||| :|| :|||||| ||: ||||| ||| : |  |||||||||||  ||||||||||||
Db     238 AEKPVTIHATPEGTSEACRMILEIMQKEADETKLAEEIPLKILAHNGLVGRLIGKEGRNL 297

Qy     300 KKIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIASMN 359
           |||| :| |||||| ||:|::|||||||||||| || || || ||||||:||::|||: ::|
Db     298 KKIEHETGTKITISSLQDLSIYNPERTITVKGTVEACASAEIEIMKKLREAFENDMLAVN 357

Qy     360 LQAHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMTPPYPQFEQ-SETETVHQFIPALSVG 418
             :          | |         :|  |     |:|      | | |: ||| :||
Db     358 THS----------GYF---------SSLYPHHQFGPFPHHHSYPEQEIVNLFIPTQAVG 397

Qy     419 AIIGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEEN 478
           ||||||:| ||||||:|||||||||||||| ||   |||||||||||||||||||:||:||||
Db     398 AIIGKKGAHIKQLARFAGASIKIAPAEGPDVSERMVIITGPPEAQFKAQGRIFGKLKEEN 457

Qy     479 FVSPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVVVK 538
           | :|||||||||||||||  |||||||||||||||||:||||:|||||||||||::|:|:
Db     458 FFNPKEEVKLEAHIRVPSSTAGRVIGKGGKTVNELQNLTSAEVIVPRDQTPDENEEVIVR 517

Qy     539 ITGHFYACQVAQRKIQEILTQVKQHQQQ 566
           | |||:| | ||||||:||: |||| :|:
Db     518 IIGHFFASQTAQRKIREIVQQVKQQEQK 545


RESULT 10
Q9VZ69
ID   Q9VZ69      PRELIMINARY;      PRT;   558 AA.
AC   Q9VZ69;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   CG1691 protein.
GN   IMP OR CG1691.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003484; AAF47958.1; -.
DR   FlyBase; FBgn0030235; Imp.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   Pfam; PF00013; KH-domain; 4.
DR   SMART; SM00322; KH; 4.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
SQ   SEQUENCE   558 AA;  61748 MW;  833C9C56954C139C CRC64;

  Query Match             28.3%;  Score 835.5;  DB 5;  Length 558;
  Best Local Similarity   38.5%;  Pred. No. 3.6e-47;
  Matches  198;  Conservative   86;  Mismatches  147;  Indels   83;  Gaps   12;

Qy     121 VVNVTYSSKDQARQALDK----LNGFQLENFTLKVAYIPDEMAAQQNPLQQPRGRRGLGQ 176
           : |  |  |    : ||:    ||| : |           :: |:|    | | :|
Db      14 ISNNYYQQKQSLIRYLDRAAVGLNGVEFEG---------SKLHAEQLDKNQRRSQR---- 60

Qy     177 RGSSRQGSPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKEN 236
             : |   ||      :  | |||:|| :: |||||||::|:||| ||:|:::::|||||||
Db      61 --NQRNPYPGMPGPGRQADFPLRILVQSEMVGAIIGRQGSTIRTITQQSRARVDVHRKEN 118

Qy     237 AGAAEKSITILSTPEGTSAACKSILEIMHKEAQDIKFTE------EIPLKILAHNNFVGR 290
           |: ||||||   ||  : ||| |||:| :||     |      || |||||||: ||| 
Db     119 VGSLEKSITIYGNPENCTNACKRILEVMQQEAISTNKGELSPECSEICLKILAHNNLIGR 178

Qy     291 LIGKEGRNLKKIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRES 350
           :||| |  :|:| ||||||||:| : ::  :| || ||||| :|  ::|| :|  |:|:|
Db     179 IIGKSGNTIKRIMQDTDTKITVSSINDINSFNLERIITVKGLIENMSRAENQISTKLRQS 238

Qy     351 YENDIASMNLQAHLIPGLNLNALGLFPPTSGMPPPTSGP-----------PSAMTPPYP 398
           ||||: :|  |: : |||:  |: :  | :||   || |           |::: ||
Db     239 YENDLQAMAPQSLMFPGLHPMAM-MSTPGNGMVFNTSMPFPSCQSFAMSKTPASVVPPV- 296

Qy     399 QFEQSETETVHQFIPALSVGAIIGKQGQHIKQLSRFAGASIKIAPAEA----PDAKVRMV 454
           |      || : :||  :|||||| :| ||: : ||: ||:|||| :|         | |
Db     297 -FPNDLQETTYLYIPNNAVGAIIGTRGSHIRSIMRFSNASLKIAPLDADKPLDQQTERKV 355

Qy     455 IITGPPEAQFKAQGRIYGKIKEENFVSPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQ 514
           | | || |:||  |: |::|| |:   ::|:|   : | |   ||:|||||: | |||
Db     356 TIVGTPEGQWKAQYMIFEKMREEGFMCGTDDVRLTVELLVASSQVGRIIGKGGQNVRELQ 415

Qy     515 NLSSAEVVVPR----------DQTPDENDQVVVKITGHFYACQVAQRKIQEIL------- 557
           :: : : :|           ::||       | | | ||: | |||:|: ::
Db     416 RVTGSVIKLPEHALAPPSGGDEETP-------VHIIGLFYSVQSAQRRIRAMMLSTNPPP 468

Qy     558 ---------TQVKQHQQQKA------LQSGPPQS 576
                   |::| ||  |        |   |||
Db     469 ITKKQKAAKEQLQQQQQSLAGAASSGSQQQQPQS 502


RESULT 11
Q9NGS9
ID   Q9NGS9      PRELIMINARY;      PRT;   566 AA.
AC   Q9NGS9;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
```

```
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   IGF-II mRNA-binding protein (SD07045p).
GN   IMP OR CG1691.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Nielsen J., Nielsen F.C., Christiansen J.;
RT   "Cloning and Expression of a Drosophila Homolog of IMP/Vg1-RBP.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA   Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA   Nunoo J., Pacleb J., Paragas V., Park S., Phouanenavong S., Wan K.,
RA   Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF241237; AAF63331.1; -.
DR   EMBL; AY069821; AAL39966.1; -.
DR   FlyBase; FBgn0030235; Imp.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   Pfam; PF00013; KH-domain; 4.
DR   SMART; SM00322; KH; 4.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
SQ   SEQUENCE   566 AA;  62129 MW;  B4C74C5C99B0C830 CRC64;

  Query Match             28.2%;  Score 835;  DB 5;  Length 566;
  Best Local Similarity   38.8%;  Pred. No. 3.9e-47;
  Matches  197;  Conservative   86;  Mismatches  147;  Indels   78;  Gaps   12;

Qy    121 VVNVTYSSKDQARQALDK----LNGFQLENFTLKVAYIPDEMAAQQNPLQQPRGRRGLGQ 176
Db     14 ISNNYYQQKQSLIRYLDRAAVGLNGVEFEG---------SKLHAEQLDKNQRRSQR---- 60

Qy    177 RGSSRQGSPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKEN 236
Db     61 --NQRNPYPGMPGPGRQADFPLRILVQSEMVGAIIGRQGSTIRTITQQSRARVDVHRKEN 118

Qy    237 AGAAEKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEG 296
Db    119 VGSLEKSITIYGNPENCTNACKRILEVMQQEAISTN-KGEICLKILAHNNLIGRIIGKSG 177

Qy    297 RNLKKIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIA 356
Db    178 NTIKRIMQDTDTKITVSSINDINSFNLERIITVKGLIENMSRAENQISTKLRQSYENDLQ 237

Qy    357 SMNLQAHLIPGLNLNALGLFPPTSGMPPPTSGP----------PSAMTPPYPQFEQSE 404
Db    238 AMAPQSLMFPGLHPMAM-MSTPGNGMVFNTSMPFPSCQSFAMSKTPASVVPPV--FPNDL 294

Qy    405 TETVHQFIPALSVGAIIGKQGQHIKQLSRFAGASIKIAPAEA----PDAKVRMVIITGPP 460
Db    295 QETTYLYIPNNAVGAIIGTRGSHIRSIMRFSNASLKIAPLDADKPLDQQTERKVTIVGTP 354

Qy    461 EAQFKAQGRIYGKIKEENFVSPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAE 520
Db    355 EGQWKAQYMIFEKMREEGFMCGTDDVRLTVELLVASSQVGRIIGKGGQNVRELQRVTGSV 414

Qy    521 VVVPR----------DQTPDENDQVVVKITGHFYACQVAQRKIQEIL------------ 557
Db    415 IKLPEHALAPPSGGDEETP-------VHIIGLFYSVQSAQRRIRAMMLSTNPPPITKKQK 467

Qy    558 ---TQVKQHQQQKA------LQSGPPQS 576
Db    468 AAKEQLQQQQQSLAGAASSGSQQQQPQS 495


RESULT 12
```

```
Q9D054
ID   Q9D054      PRELIMINARY;      PRT;   100 AA.
AC   Q9D054;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   2610101N11Rik protein.
GN   IGF2BP3 OR 2610101N11RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=EMBRYO;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK011797; BAB27848.1; -.
DR   MGD; MGI:1890359; Igf2bp3.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00076; rrm; 1.
DR   SMART; SM00360; RRM; 1.
DR   PROSITE; PS50102; RRM; 1.
SQ   SEQUENCE   100 AA;  11249 MW;  4D871E37EB9D5466 CRC64;

  Query Match             16.0%;  Score 473;  DB 11;  Length 100;
  Best Local Similarity   88.7%;  Pred. No. 3.1e-24;
  Matches   86;  Conservative    9;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGK 60
Db      1 MNKLYIGNLSDHAGPADLESVFKDAKIPVAGPFLVKTGYAFVDCPDEGWALKAIEALSGK 60

Qy     61 IELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVL 97
Db     61 MELHGKPMEVEHSVPKRQRIRKLQIRNIPPHLQWEMM 97


RESULT 13
Q21605
ID   Q21605      PRELIMINARY;      PRT;   854 AA.
AC   Q21605;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   M88.5 protein.
GN   M88.5.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Sulston J.E.;
```

```
RL   Submitted (JUN-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
DR   EMBL; Z34802; CAA84338.1; -.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   Pfam; PF00013; KH-domain; 4.
DR   SMART; SM00322; KH; 4.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
SQ   SEQUENCE   854 AA;  94553 MW;  D3B40412353ECC21 CRC64;

  Query Match          14.2%;  Score 418.5;  DB 5;  Length 854;
  Best Local Similarity   24.1%;  Pred. No. 2.6e-19;
  Matches  160;  Conservative  103;  Mismatches  215;  Indels  187;  Gaps   27;

Qy      78 QRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETAVVN--------VTYSSK 129
           |:::: | |   |   : | ::  |       :|::   : | :|         |
Db     167 QQMQQQQQSNQSAHFMHQQLQAVQQQ-----QAQQMHHRLQGAPINPQQFMVPPPTMMQP 221

Qy     130 DQARQALDKLNGFQLENFTLKVAYIPDEMA--AQQ--NPLQQPRGRRGLGQRGSSRQGSP 185
             | :||  :    |:    :      |  |   |||  :| || :  :  ||   |   |
Db     222 QQMQQA-QQQQAQQMHQMQIHHQQHPQMMQQHAQQGYHPHQQNQQHQA-GQHQQSHHQSQ 279

Qy     186 G-------SVSKQKP----------C---DLPLRLLVPTQFVGAIIGKEGATIRNITKQ 224
                 : |   |            |   | |:| :|  ::   |||  |:||::|
Db     280 NHNQHRNHNQSHSGPHHIPQNLMMPRCMLKDWPIRCVVEGKYHAVIIGPNGSTIKDIASS 339

Qy     225 TQSKID---VHRKENA--GAAEKSITILSTPEGTSAACKSILEIMHKEA--QDIKFTEEI 277
           |: ::|   : :||    |  :: :|:    |  : |   ||::: ||   |:    :
Db     340 TRCRVDFVNLSKKERTVLGNNDRILTVHGVAEQATKAVARILDVIQSEAVKDDVNVGADT 399

Qy     278 PLKILAHNNFVGRLIGKEGRNLKKIEQDTDTKITISPLQE-------LT----LYNPERT 326
           |:: |||   ||||| | ::|:| | | | ||::   |       ||    |   |||
Db     400 VLRMRAHNQLCGRLIGKAGSSIKEIMQKTGTNITVTKYIEPPGGISGLTANELLGLMERT 459

Qy     327 ITVKG-NVETCAKAEEEIMKKIRESYENDIASMNLQAHLIPGLNLNALGLFPPTSGMPP- 384
           | |:| ::|   :||  |  |::: ||:| : :  |:   |          ||   |||
Db     460 IMVRGPSIEAVVQAEALISAKLKKCYESD-SQLRAQSMQCP---------MPPMM-MPPI 508

Qy     385 -PTSGPPSAMTPPY----------PQFEQSETETVHQ---------------------- 410
           |    ||:: |:           |   |    |||
Db     509 LPPGASSSAVSAPHFIPTPVGVMQVQHFASSQHLVHQNANNSFLQPGVLQIQPGTTNLRQ 568

Qy     411 ---FIPALSVGAIIGKQGQHIKQLSRFAGASIKIAPAEAPDAKV---------------- 451
              ::|   :||:|| :|::|| : |  |||:||   |||: | .
Db     569 VRMWVPDSMIGALIGAKGKNIKMIIRDTGASVKI---EAPEEKTQREAEEAEKKRKLDET 625

Qy     452 ------------------------------------RMVIITGPPEAQFKAQGRIYGK 473
                                                ||| | |      |||  :: |
Db     626 DSGCEGVASGDHPQEFLEDNATINSSDAIEEKPKPVSERMVTINGDDLQLLKAQSYVFSK 685

Qy     474 IKEENFVSP--------KEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPR 525
           | | :   |          :::   : ||:   ||:|||||: | ||| :: | | :|
Db     686 IAETSSSLPSSGMDGDRSHMLRIRTEVSVPTRIIGRIIGKGGQNVRELQRITGAVVKIPE 745

Qy     526 DQTPD----------ENDQVVVKITGHFYACQVAQRKIQEILTQVKQ---HQQQKA-LQS 571
           ::           | |  :::  |: |:    | ::  :: :  :   |: : :  :
Db     746 EERNGGEVYRHDDGLEEDMTMIRTIGNMYSTHNVQFRLAHLVNEYYRSGDHRNKSSDYKG 805

Qy     572 GPPQS 576
           | | |
Db     806 GRPHS 810


RESULT 14
Q9LI28
ID   Q9LI28      PRELIMINARY;      PRT;   774 AA.
AC   Q9LI28;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   ESTs D23839(R0339).
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. NIPPONBARE;
RA   Sasaki T., Matsumoto T., Yamamoto K.;
RT   "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT   clone:P0708G02.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP001539; BAA92910.1; -.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   Pfam; PF00013; KH-domain; 4.
DR   SMART; SM00322; KH; 4.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
SQ   SEQUENCE   774 AA;  84066 MW;  5992CCFE0F74BB56 CRC64;

  Query Match          8.9%;  Score 262.5;  DB 10;  Length 774;
  Best Local Similarity   21.8%;  Pred. No. 4.8e-09;
  Matches  129;  Conservative   73;  Mismatches  185;  Indels  205;  Gaps   20;

Qy     156 DEMAAQQNPLQQPRGRRGLGQRGSSRQGSPGSVSKQKPCDLP----------LRLLVP 203
           ::||::  | ::|       |: || |   |   | |   |             |:| |
Db      57 EDMASRSTPSKRP------FQKNSSEQNGRGKWQKTKHNSLQQPQLIVQPGVPIFRILCP 110

Qy     204 TQFVGAIIGKEGATIRNITKQTQSKI-------------------------DVHRKENA 237
           |   | :||| |  |  | ::|  ||                        | |||
Db     111 TSKSGNVIGKGGGIIAKIRQETGVKIRVDEVVPGCDERVIVITAIDKDREVSHVQTKEND 170

Qy     238 GA----------------------------AEKSI---------------------- 244
           |                            :|| :
Db     171 GGVACSVDGNHGMEKDHTKEEKDESNKEMDDSEKGLGKEEKDDLEKDHDKEDRDESGKDN 230

Qy     245 -----------------TILSTPEGTSAACKSILEIMHKEAQDIKFTEE---------- 276
                           |   :|   | |:|| :      |  |  |
Db     231 DKKADDSSVAKDTNSEPEAQLELEKGMPLAVKAILLVF-----DRIFVNEMENGTGDASG 285

Qy     277 ----IPLKILAHNNFVGRLIGKEGRNLKKIEQDTDTKITISPLQELTLYNPERTITVKGN 332
               : |:: |  :: || |:|| |  :|::  |:  :| :|              |
Db     286 ERNHVSLRLLVLDSQVGWLLGKNGSVIKQMSTDSCCEIRVS---------------KDK 329

Qy     333 VETCAKAEEEIMKKIRE---SYENDIASMNLQAHL---IPGLNLNALGLF--PPTS---- 380
           :  ||   :|: : : :   :|   |: : :   |     |||  |  ||  ||
Db     330 LPLCALPRDELCQILHQKIITYSLHISFLQITGELDSVRKGLNTVAQLLFTHPPKESDVL 389

Qy     381 -------------------GMPP----PTSGPPSAMTPPYPQ-------FEQSETETVH 409
                             || |     |  ||  |  | :|:           |  |
Db     390 GAHNSGSSRSFFNQPDVLPPGMQPNLHLPFQGPNVAHLPNFPEALMHGHGSVPPEPLTFR 449

Qy     410 QFIPALSVGAIIGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPP---EAQFKA 466
               :  || |||| | :|| :    |  ||:     | :: |:| |:||    :   |
Db     450 LLCSSDKVGGIIGKGGNNIKSIQNDTGCEIKVLDT-VPKSEDRIVFISGPAHPGDGISPA 508

Qy     467 QGRIYGKIKEENFVSPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVV-PR 525
           |  |    ::   |  :|    : |    | ::|||  : |:: || | ::|  :
Db     509 QNAILHVQRKIVPTSNTKEGPAICRLIVSPNQVGCLLGKGGSIIAEMRKLSGAHIIVLSK 568

Qy     526 DQTP---DENDQVVVKITGHFYACQVAQRKIQEILTQVKQHQQQKALQSGPP 574
           |: |    |||: ||:|:|   | :  |  : :|  ::: |  :  : |  |
Db     569 DKIPKGVPENDE-VVQISG---ASEAIQEALMQITARLRNHLFRDRMASTVP 616


RESULT 15
Q8S7G1
ID   Q8S7G1      PRELIMINARY;      PRT;   542 AA.
AC   Q8S7G1;
```

```
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative nucleic acid binding protein.
GN   OSJNBB0048D20.13.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. NIPPONBARE;
RA   Buell C.R., Yuan Q., Ouyang S., Liu J., Moffat K.S., Hill J.N.,
RA   Gansberger K., Brenner M., Burgess S., Hance M., Shvartsbeyn M.,
RA   Tsitrin T., Riggs F., Hsiao J., Zismann V., Blunt S., Pai G.,
RA   VanAken S.E., Utterback T.R., Feldblyum T.V., Kalb E., Quackenbush J.,
RA   Salzberg S.L., White O., Fraser C.M.;
RT   "Oryza sativa chromosome 3 BAC OSJNBb0048D20 genomic sequence.";
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC084766; AAL82529.1; -.
SQ   SEQUENCE   542 AA;  57843 MW;  B28A54BB2450F336 CRC64;

  Query Match             8.8%;  Score 260.5;  DB 10;  Length 542;
  Best Local Similarity   22.5%;  Pred. No. 3.9e-09;
  Matches  105;  Conservative   94;  Mismatches  191;  Indels   77;  Gaps   17;

Qy    168 PRGRRGLGQR-------GSSRQGSPG-SVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIR 219
          |  |   |:|       |  :: :||       | |   | | |:: :|:|||: |   :
Db      4 PGHRNSHGKRHSDYAENGGGKRRNPGDDTYAPGPDDTVYRYLCPSRKIGSIIGRGGEIAK 63

Qy    220 NITKQTQSKIDVHRKENAGAAEKSITILSTPEGTSAACKSILEIMHK--EAQDIKF---- 273
           :   ||:|| :  :  :|  |: ||| |:   |:    ::::  |   ||| |
Db     64 QMRADTQAKIRI-GESVSGCDERVITIFSSSRETN----TLVDAEDKVCPAQDALFRVHE 118

Qy    274 ----------------TEEIPLKILAHNNFVGRLIGKEGRNLKKIEQDTDTKITISPLQE 317
                          :: :::|  :: :| :||| |  :: |  ||   | :   :
Db    119 KLSIDDDIGNEESDEGLAQVTVRLLVPSDQIGCIIGKGGHIIQGIRSDTGAHIRVLSNEN 178

Qy    318 L--TLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIASMNLQA----------HL 364
          |     : :  : : |:     ||  ::  ::   ::|   | :| |           ||
Db    179 LPACAISGDELLQISGDSTVVRKALLQVSSRL---HDNPSRSQHLLASSMTQPYPVGSHL 235

Qy    365 IPGLNLNALGLFPPTS---GMPPPTSGPPSAMTPPYPQFEQSETETVHQFIPALSVGAII 421
                :|: |  |   |     :|   ::  |  :   ::  ::     | :|| :|
Db    236 GSSSTAPVVGITPLISSYGGYKGDVAGDWPSIYQPRREESSAKEFSLRLLCAASNVGGVI 295

Qy    422 GKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEENFVS 481
          || |  |||: : :|| ||:  :   |      |||   :  |  :  :  |     :
Db    296 GKGGGIIKQIRQESGAFIKVDSSNTEDD----CIITVSAKEFF--EDPVSPTINAAVHLQ 349

Qy    482 PKEEVKLEAHIRVPSFAA---------GRVIGKGGKTVNELQNLSSAEV-VVPRDQTPD- 530
          |:   | :    :||:           | :|||||  : |::  | | : :: ::  |
Db    350 PRCSEKTDPESAIPSYTTRLLVSTSRIGCLIGKGGSIITEIRRTSRANIRILSKENVPKV 409

Qy    531 -ENDQVVVKITGHFYACQVAQRKIQEILTQVKQH--QQQKALQSGPP 574
           |: :|:|:|      | :  : :| |::| :  ::: ||   ||
Db    410 AAEDEEMVQISGDL---DVVRHALLQITTRLKANFFEREGALSGFPP 453


Search completed: April 24, 2003, 16:14:14
Job time : 42 secs
```

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: April 24, 2003, 16:11:06 ; Search time 14 Seconds
(without alignments)
1715.342 Million cell updates/sec

Title: US-09-897-778-176
Perfect score: 2956
Sequence: 1 MNKLYIGNLSENAAPSDLES.....VKQHQQQKALQSGPPQSRRK 579

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	227	7.7	510	1	NOA1_HUMAN	P51513 homo sapien
2	224.5	7.6	339	1	PCB3_HUMAN	P57721 homo sapien
3	219.5	7.4	339	1	PCB3_MOUSE	P57722 mus musculu
4	213.5	7.2	403	1	PCB4_HUMAN	P57723 homo sapien
5	212.5	7.2	403	1	PCB4_MOUSE	P57724 mus musculu
6	210.5	7.1	606	1	Y475_ARATH	P58223 arabidopsis
7	209.5	7.1	356	1	PCB1_HUMAN	Q15365 homo sapien
8	209.5	7.1	356	1	PCB1_RABIT	O19048 oryctolagus
9	207.5	7.0	1268	1	VGLN_HUMAN	Q00341 homo sapien
10	204	6.9	413	1	PBP2_YEAST	P38151 saccharomyc
11	202.5	6.9	365	1	PCB2_HUMAN	Q15366 homo sapien
12	202	6.8	362	1	PCB2_MOUSE	Q61990 mus musculu
13	194.5	6.6	1270	1	VGLN_CHICK	P81021 gallus gall
14	190	6.4	463	1	ROK_HUMAN	Q07244 homo sapien
15	190	6.4	463	1	ROK_RABIT	O19049 oryctolagus
16	190	6.4	464	1	ROK_MOUSE	Q60577 mus musculu
17	183	6.2	1222	1	S160_YEAST	P06105 saccharomyc
18	170.5	5.8	629	1	PAB2_ARATH	P42731 arabidopsis
19	164.5	5.6	644	1	PAB4_HUMAN	Q13310 homo sapien
20	160.5	5.4	636	1	PAB1_MOUSE	P29341 mus musculu
21	159.5	5.4	470	1	NR54_HUMAN	Q15233 homo sapien
22	159.5	5.4	636	1	PAB1_HUMAN	P11940 homo sapien
23	156	5.3	633	1	PABP_XENLA	P20965 xenopus lae
24	152.5	5.2	381	1	YBD2_YEAST	P38199 saccharomyc
25	150	5.1	344	1	SFR6_HUMAN	Q13247 homo sapien
26	147.5	5.0	747	1	FCA_ARATH	O04425 arabidopsis
27	145.5	4.9	414	1	NSR1_YEAST	P27476 saccharomyc
28	145.5	4.9	489	1	SFR4_MOUSE	Q8ve97 mus musculu
29	145	4.9	414	1	NOP3_YEAST	Q01560 saccharomyc
30	139.5	4.7	522	1	PAB2_HUMAN	Q15097 homo sapien
31	138.5	4.7	424	1	S3B4_HUMAN	Q15427 homo sapien
32	135.5	4.6	359	1	ELV2_HUMAN	Q12926 homo sapien
33	135.5	4.6	360	1	ELV2_MOUSE	Q60899 mus musculu
34	135	4.6	367	1	ELV3_HUMAN	Q14576 homo sapien
35	135	4.6	367	1	ELV3_MOUSE	Q60900 mus musculu
36	135	4.6	494	1	SFR4_HUMAN	Q08170 homo sapien
37	134	4.5	285	1	CABA_MOUSE	Q99020 mus musculu
38	134	4.5	307	1	SXL_MEGSC	O01671 megaselia s
39	133	4.5	373	1	ELV4_RAT	O09032 rattus norv
40	133	4.5	380	1	ELV4_HUMAN	P26378 homo sapien
41	133	4.5	385	1	ELV4_MOUSE	Q61701 mus musculu
42	130	4.4	633	1	ROR_HUMAN	O43390 homo sapien
43	129.5	4.4	6359	1	BACC_BACLI	O68008 b bacitraci
44	129	4.4	653	1	PABP_SCHPO	P31209 schizosacch
45	128.5	4.3	660	1	PAB3_ARATH	O64380 arabidopsis

ALIGNMENTS

RESULT 1
NOA1_HUMAN
ID NOA1_HUMAN STANDARD; PRT; 510 AA.
AC P51513;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Onconeural ventral antigen-1 (NOVA-1) (Paraneoplastic Ri antigen)
DE (Ventral neuron-specific protein 1).
GN NOVA1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cerebellum, and Hippocampus;
RX MEDLINE=94000830; PubMed=8398153;
RA Buckanovich R.J., Posner J.B., Darnell R.B.;
RT "Nova, the paraneoplastic Ri antigen, is homologous to an RNA-binding
RT protein and is specifically expressed in the developing motor
RT system.";
RL Neuron 11:657-672(1993).
RN [2]
RP SEQUENCE OF 1-34 FROM N.A.
RC TISSUE=Fetal brain;
RA Dmitrenko V.V., Garifulin O.M., Shostak K.A., Smikodub A.I.,
RA Kavsan V.M.;
RL Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.
CC -!- FUNCTION: MAY REGULATE RNA SPLICING OR METABOLISM IN A SPECIFIC
CC SUBSET OF DEVELOPING NEURONS.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: BRAIN.
CC -!- DISEASE: ANTIBODIES ARE SEEN AGAINST THIS ANTIGEN IN THE PATIENTS
CC SUFFERING WITH PARANEOPLASTIC OPSOCLONUS-ATAXIA (POA). POA IS A
CC DISORDER IN WHICH ABNORMAL MOTOR CONTROL OF THE EYES, TRUNK AND
CC LIMBS DEVELOP IN WOMEN WITH BREAST OR SMALL LUNG CANCER. POA
CC PATIENTS SUFFER FROM OPSOCLONUS, A CHAOTIC EYE MOVEMENT DISORDER
CC ATTRIBUTABLE TO A LACK OF INHIBITION OF BURST NEURONS IN THE BRAIN
CC STEM; MYOCLONUS, A DISORDER PERHAPS ATTRIBUTING TO A FAILURE OF
CC INHIBITORY CONTROL OVER SPINAL MOTOR NEURONS; AND TRUNCAL ATAXIA,
CC REFERABLE TO CEREBELLAR DYSFUNCTION; THERE ARE NO CORTICAL OR
CC SENSORY DEFICITS.
CC -!- SIMILARITY: CONTAINS 3 KH DOMAINS.
CC ---

CC ---
DR EMBL; U04840; AAA16022.1; -.
DR EMBL; Z70771; CAA94810.1; -.
DR Genew; HGNC:7886; NOVA1.

```
DR   MIM; 602157; -.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   Pfam; PF00013; KH-domain; 3.
DR   SMART; SM00322; KH; 3.
DR   PROSITE; PS50084; KH_TYPE_1; 3.
KW   Nuclear protein; RNA-binding; Repeat; Alternative splicing; Antigen.
FT   DOMAIN       52     87       KH 1.
FT   DOMAIN      177    212       KH 2.
FT   DOMAIN      427    462       KH 3.
FT   VARSPLIC     88     90       MISSING (IN TUMOR ISOFORM).
FT   VARSPLIC    153    176       MISSING (IN ISOFORM 3).
FT   VARSPLIC    177    184       VKIIVPNS -> KHNISWIS (IN TUMOR ISOFORM).
FT   VARSPLIC    185    510       MISSING (IN TUMOR ISOFORM).
SQ   SEQUENCE   510 AA;  52055 MW;  7B9BA8B67CAB6984 CRC64;

  Query Match             7.7%;  Score 227;  DB 1;  Length 510;
  Best Local Similarity   22.4%;  Pred. No. 5.1e-07;
  Matches  119;  Conservative   77;  Mismatches  179;  Indels  156;  Gaps   20;

Qy    155 PDEMAAQQNPLQQPRGRRGLGQRGSSRQGSPGSVSKQKPCDLPLRLLVPTQFVGAIIGKE 214
Db     23 PD---SRKRPLEAP------PEAGSTKRTNTGEDGQYF-----LKVLIPSYAAGSIIGKG 68

Qy    215 GATIRNITKQTQSKIDVHRKENA-----GAAEKSITILSTPEGTSAACKSILEIMHKEAQ 269
Db     69 GQTIVQLQKETGATIKLSKLSKSKDFYPGTTERVCLIQGTVEALNAVHGFIAEKIREMPQ 128

Qy    270 DIKFTEEIPL-----------------------------------------KILAHNNFVGR 290
Db    129 NVAKTEPVSILQPQTTVNPDRIKQTLPSSPTTTKSSPSDPMTTSRANQVKIIVPNSTAGL 188

Qy    291 LIGKEGRNLKKIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRE- 349
Db    189 IIGKGGATVKAVMEQSGAWVQLSQKPD-GINLQERVVTVSGEPEQNRKAVELIIQKIQED 247

Qy    350 ---------SYEN---DIASMN----------------LQAHLIPGLNLNALGLFPPT- 379
Db    248 PQSGSCLNISYANVTGPVANSNPTGSPYANTAEVLPTAAAAAGLLGHANLAGVAAFPAVL 307

Qy    380 SGMPPPTSGPPSAMTPPYPQFEQSETETVHQFIPALSVGAIIGKQGQHIKQLSRFAGASI 439
Db    308 SGF---TGNDLVAIT--------SALNTL------ASYGYNLNTLGLGLSQAAA-TGALA 349

Qy    440 KIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKI---------------------------- 474
Db    350 AAAASANPAAAAANLLATYASEA--SASGSTAGGTAGTFALGSLAAATAATNGYFGAASP 407

Qy    475 --------KEENFVSPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPR- 525
Db    408 LAASAILGTEKSTDGSKDVVE----IAVPENLVGAILGKGGKTLVEYQELTGARIQISKK 463

Qy    526 -DQTPDENDQVVVKITGHFYACQVAQRKIQEILTQVKQHQQQKALQSGPPQ 575
Db    464 GEFVPGTRNRKVT-ITGTPAATQAAQYLITQRIT------YEQGVRAANPQ 507


RESULT 2
PCB3_HUMAN
ID   PCB3_HUMAN     STANDARD;      PRT;   339 AA.
AC   P57721;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Poly(rC)-binding protein 3 (Alpha-CP3).
GN   PCBP3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20396135; PubMed=10936052;
RA   Makeyev A.V., Liebhaber S.A.;
RT   "Identification of two novel mammalian genes establishes a subfamily
RT   of KH-domain RNA-binding proteins.";
RL   Genomics 67:301-316(2000).
CC   -!- FUNCTION: SINGLE-STRANDED NUCLEIC ACID BINDING PROTEIN THAT
CC       BINDS PREFERENTIALLY TO OLIGO DC (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- SIMILARITY: CONTAINS 3 KH DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF176329; AAG09240.1; -.
DR   HSSP; Q07244; 1KHM.
DR   Genew; HGNC:8651; PCBP3.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   Pfam; PF00013; KH-domain; 3.
DR   SMART; SM00322; KH; 3.
DR   PROSITE; PS50084; KH_TYPE_1; 3.
KW   Nuclear protein; RNA-binding; Ribonucleoprotein; DNA-binding;
KW   Repeat.
FT   DOMAIN       13     63       KH 1.
FT   DOMAIN       97    150       KH 2.
FT   DOMAIN      261    313       KH 3.
SQ   SEQUENCE   339 AA;  35938 MW;  F9F19FF7590C4188 CRC64;

  Query Match             7.6%;  Score 224.5;  DB 1;  Length 339;
  Best Local Similarity   25.0%;  Pred. No. 4.1e-07;
  Matches   83;  Conservative   65;  Mismatches  119;  Indels   65;  Gaps   14;

Qy    196 LPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAAEKSITILSTPEGTSA 255
Db     14 LTIRLLMHGKEVGSIIGKKGETVKKMREESGARINI----SEGNCPERIVTITGP--TDA 67

Qy    256 ACKSILEIMHKEAQDI-------KFTEEIP--LKILAHNNFVGRLIGKEGRNLKKIEQDT 306
Db     68 IFKAFAMIAYKFEEDIINSMSNSPATSKPPVTLRLVVPASQCGSLIGKGGSKIKEIREST 127

Qy    307 DTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRES--------YENDIASM 358
Db    128 GAQVQVA--GDMLPNSTERAVTISGTPDAIIQCVKQICVVMLESPPKGATIPYRPKPAST 185

Qy    359 NL-----QAHLIPG----------LNLNALGL----FPPTSGMPPPTSGPPSAMTPPYPQ 399
Db    186 PVIFAGGQAYTIQGQYAIPHPDQLTKLHQLAMQQTPFPPLGQTNPAF---PGEKLPLHSS 242

Qy    400 ------------FEQSETETVHQF-IPALSVGAIIGKQGQHIKQLSRFAGASIKIAPAEA 446
Db    243 EEAQNLMGQSSGLDASPPASTHELTIPNDLIGCIIGRQGTKINEIRQMSGAQIKIANATE 302

Qy    447 PDAKVRMVIITGPPE----AQFKAQGRIYGKI 474
Db    303 GSSE-RQITITGTPANISLAQYLINARLTSEV 333


RESULT 3
PCB3_MOUSE
ID   PCB3_MOUSE     STANDARD;      PRT;   339 AA.
AC   P57722;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Poly(rC)-binding protein 3 (Alpha-CP3).
GN   PCBP3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
```

```
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20396135; PubMed=10936052;
RA   Makeyev A.V., Liebhaber S.A.;
RT   "Identification of two novel mammalian genes establishes a subfamily
RT   of KH-domain RNA-binding proteins.";
RL   Genomics 67:301-316(2000).
CC   -!- FUNCTION: SINGLE-STRANDED NUCLEIC ACID BINDING PROTEIN THAT
CC       BINDS PREFERENTIALLY TO OLIGO DC (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- TISSUE SPECIFICITY: UBIQUITOUS.
CC   -!- SIMILARITY: CONTAINS 3 KH DOMAINS.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AF176327; AAG09238.1; -.
DR   HSSP; Q07244; 1KHM.
DR   MGD; MGI:1890470; Pcbp3.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   Pfam; PF00013; KH-domain; 3.
DR   SMART; SM00322; KH; 3.
DR   PROSITE; PS50084; KH_TYPE_1; 3.
KW   Nuclear protein; RNA-binding; Ribonucleoprotein; DNA-binding;
KW   Repeat.
FT   DOMAIN       13     63       KH 1.
FT   DOMAIN       97    150       KH 2.
FT   DOMAIN      261    313       KH 3.
SQ   SEQUENCE   339 AA;  35958 MW;  A6F1C7C176A64F9C CRC64;

  Query Match             7.4%;  Score 219.5;  DB 1;  Length 339;
  Best Local Similarity   20.6%;  Pred. No. 8.4e-07;
  Matches   75;  Conservative   52;  Mismatches  108;  Indels  129;  Gaps    7;

Qy    277 IPLKILAHNNFVGRLIGKEGRNLKKIEQDTDTKITISPLQELTLYNPERTITVKGNVETC 336
          : :::| |   || :|||:|  :||: :::  :| ||         ||| :|: |  :
Db     14 LTIRLLMHGKEVGSIIGKKGETVKKMREESGARINISEGN-----CPERIVTITGPTDAI 68

Qy    337 AKAEEEIMKKIRESYENDIASMNLQAHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMTPP 396
           ||   |  | |   | :::                          | || ||
Db     69 FKAFAMIAYKFEEDIINSMSN------------------------SPATSKPP------ 97

Qy    397 YPQFEQSETETVHQFIPALSVGAIIGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVII 456
                 |:   :||   |::||| |  ||::    || :::|    |::  | | |
Db     98 ---------VTLRLVVPASQCGSLIGKGGSKIKEIRESTGAQVQVAGDMLPNSTERAVTI 148

Qy    457 TGPPEAQFK-------------------------------AQGRIY------------ 471
          :| |:| |:                               | |: |
Db    149 SGTPDAIFQCVKQICVVMLESPPKGATIPYRPKPACTPVIFAGGQAYTIQGQYAIPHPDQ 208

Qy    472 --------------GKIKEENFVSPKEEVKLEA----------------------HIRV 494
                        : : |   | |:: | :                      : :
Db    209 LTKLHQLAMQQTPFPPLGQTNPAFPGEKLPLHSSEEAQNLMGQSSGLDASPPASTHELTI 268

Qy    495 PSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVVVKITGHFYACQVAQRKIQ 554
          |:   | :||: |  :||:: :| |:: :          |:   |||      :||  |
Db    269 PNDLIGCIIGRQGTKINEIRQMSGAQIKIANATEGSSERQIT--ITGTPANISLAQYLIN 326

Qy    555 EILT 558
            ||
Db    327 ARLT 330


RESULT 4
PCB4_HUMAN
ID   PCB4_HUMAN     STANDARD;      PRT;   403 AA.
```

```
AC   P57723;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Poly(rC)-binding protein 4 (Alpha-CP4).
GN   PCBP4.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20396135; PubMed=10936052;
RA   Makeyev A.V., Liebhaber S.A.;
RT   "Identification of two novel mammalian genes establishes a subfamily
RT   of KH-domain RNA-binding proteins.";
RL   Genomics 67:301-316(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA   Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA   Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA   Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA   Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA   Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA   Ninomiya K., Iwayanagi T.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lung;
RA   Strausberg R.;
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: SINGLE-STRANDED NUCLEIC ACID BINDING PROTEIN THAT
CC       BINDS PREFERENTIALLY TO OLIGO DC (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- SIMILARITY: CONTAINS 3 KH DOMAINS.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AF176330; AAG09241.1; -.
DR   EMBL; AK023993; BAB14761.1; -.
DR   EMBL; BC003008; AAH03008.1; -.
DR   EMBL; BC004153; AAH04153.1; -.
DR   Genew; HGNC:8652; PCBP4.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   Pfam; PF00013; KH-domain; 3.
DR   SMART; SM00322; KH; 3.
DR   PROSITE; PS50084; KH_TYPE_1; 3.
KW   Nuclear protein; RNA-binding; Ribonucleoprotein; DNA-binding;
KW   Repeat.
FT   DOMAIN       17     67       KH 1.
FT   DOMAIN      101    154       KH 2.
FT   DOMAIN      241    293       KH 3.
SQ   SEQUENCE   403 AA;  41481 MW;  3D99F762A9471265 CRC64;

  Query Match             7.2%;  Score 213.5;  DB 1;  Length 403;
  Best Local Similarity   25.6%;  Pred. No. 2.5e-06;
  Matches   81;  Conservative   61;  Mismatches  126;  Indels   49;  Gaps   13;

Qy    183 GSPGSVSKQKPCD--LPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAA 240
          || | : ::      | ||:|:  : ||:||||:| |:: | :|: ::| :    : |:
Db      3 GSDGGLEEEPELSITLTLRMLMHGKEVGSIIGKKGETVKRIREQSSARITI----SEGSC 58

Qy    241 EKSITILSTPEGTSAACKSILEIMHKEAQDI---------KFTEEIPLKILAHNNFVGRL 291
          : || ::    |:|   ::  |  |  :|:             : |:::   :  | |
```

```
Db     59 PERITTIT--GSTAAVFHAVSMIAFKLDEDLCAAPANGGNVSRPPVTLRLVIPASQCGSL 116

Qy    292 IGKEGRNLKKIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESY 351
          ||| |  :|:|  : |  :: ::  :|  : || :|| |  :     :|    | ||
Db    117 IGKAGTKIKEIRETTGAQVQVA--GDLLPNSTERAVTVSGVPDAIILCVRQICAVILESP 174

Qy    352 ENDIASMNLQAHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMTP----------PY 397
              :  |  | |:|  : |     |       |  |:||           |:
Db    175 PK---GATIPYH--PSLSLGTV-LLSANQGF--SVQGQYGAVTPAEVTKLQQLSSHAVPF 226

Qy    398 ------PQFEQ-SETETVHQFIPALSVGAIIGKQGQHIKQLSRFAGASIKIAPAEAPDAK 450
                 |  :  ::| :    :|   :| :||:||  | :: : :|| |||   :|  |
Db    227 ATPSVVPGLDPGTQTSSQEFLVPNDLIGCVIGRQGSKISEIRQMSGAHIKIG-NQAEGAG 285

Qy    451 VRMVIITGPPEAQFKAQ 467
          | | ||| | :   ||
Db    286 ERHVTITGSPVSIALAQ 302


RESULT 5
PCB4_MOUSE
ID   PCB4_MOUSE     STANDARD;      PRT;   403 AA.
AC   P57724;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Poly(rC)-binding protein 4 (Alpha-CP4).
GN   PCBP4.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20396135; PubMed=10936052;
RA   Makeyev A.V., Liebhaber S.A.;
RT   "Identification of two novel mammalian genes establishes a subfamily
RT   of KH-domain RNA-binding proteins.";
RL   Genomics 67:301-316(2000).
CC   -!- FUNCTION: SINGLE-STRANDED NUCLEIC ACID BINDING PROTEIN THAT
CC       BINDS PREFERENTIALLY TO OLIGO DC (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- TISSUE SPECIFICITY: UBIQUITOUS.
CC   -!- SIMILARITY: CONTAINS 3 KH DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF176328; AAG09239.1; -.
DR   MGD; MGI:1890471; Pcbp4.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   Pfam; PF00013; KH-domain; 3.
DR   SMART; SM00322; KH; 3.
DR   PROSITE; PS50084; KH_TYPE_1; 3.
KW   Nuclear protein; RNA-binding; Ribonucleoprotein; DNA-binding;
KW   Repeat.
FT   DOMAIN       17     67       KH 1.
FT   DOMAIN      101    154       KH 2.
FT   DOMAIN      241    293       KH 3.
SQ   SEQUENCE   403 AA;  41398 MW;  D4DBFF490CEFE90A CRC64;

  Query Match             7.2%;  Score 212.5;  DB 1;  Length 403;
  Best Local Similarity   26.2%;  Pred. No. 2.9e-06;
  Matches   79;  Conservative   57;  Mismatches  119;  Indels   47;  Gaps   12;

Qy    196 LPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAAEKSITILSTPEGTSA 255
          | ||:|:  : ||:||||:| |:: | :|: ::| :    : |:  : || ::    |:|

Db     18 LTLRMLMHGKEVGSIIGKKGETVKRIREQSSARITI----SEGSCPERITTIT--GSTAA 71

Qy    256 ACKSILEIMHKEAQDI---------KFTEEIPLKILAHNNFVGRLIGKEGRNLKKIEQDT 306
             ::  |  | :|:             : |:::   :  | |||| |  :|:| : |
Db     72 VFHAVSMIAFKLDEDLCAAPANGGSVSRPPVTLRLVIPASQCGSLIGKAGTKIKEIRETT 131

Qy    307 DTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIASMNLQAHLIP 366
             :: ::   :|   : || :|| |  :      :|   | ||       :  |  |
Db    132 GAQVQVA--GDLLPNSTERAVTVSGVPDAIILCVRQICAVILESPPK---GATIPYH--P 184

Qy    367 GLNLNALGLFPPTSGMPPPTSGPPSAMTP----------PY------PQFEQSETE 406
           |:|  : |  . |      |   |:||             |:      |  : |
Db    185 SLSLGTV-LLSANQGF--SVQGQYGAVTPAEVTKLQQLSGHAVPFASPSVVPGMDPSTQT 241

Qy    407 TVHQF-IPALSVGAIIGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFK 465
          :  :| :|   :| :||:||  | :: : :|| |||   :|  |  | | ||| | :
Db    242 SSQEFLVPNDLIGCVIGRQGSKISEIRQMSGAHIKIG-NQAEGAGERHVTITGSPVSIAL 300

Qy    466 AQ 467
          ||
Db    301 AQ 302


RESULT 6
Y475_ARATH
ID   Y475_ARATH     STANDARD;      PRT;   606 AA.
AC   P58223; O49507;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Putative nucleic acid binding protein At4g18375.
GN   AT4G18375 OR F28J12.30.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=20083488; PubMed=10617198;
RA   Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
RA   Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
RA   Harris B., Ansorge W., Brandt P., Grivell L., Rieger M.,
RA   Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,
RA   Kreis M., Delseny M., Puigdomenech P., Watson M., Schmidtheini T.,
RA   Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
RA   Vos P., Hoheisel J., Zimmermann W., Wedler H., Ridley P.,
RA   Langham S.-A., McCullagh B., Bilham L., Robben J.,
RA   Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,
RA   Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
RA   Weitzenegger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
RA   Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,
RA   Mooijman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
RA   Berneiser S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
RA   De Keyser A., Buysshaert C., Gielen J., Villarroel R., De Clercq R.,
RA   Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
RA   Clark L., Doggett J., Hall S., Kay M., Lennard N., McLay K., Mayes R.,
RA   Pettett A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,
RA   Borkova D., Bloecker H., Scharfe M., Grimm M., Loehnert T.-H.,
RA   Dose S., de Haan M., Maarse A., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzl A.,
RA   Neumann S., Argiriou A., Vitale D., Liguori R., Piravandi E.,
RA   Massenet O., Quigley F., Clabauld G., Muendlein A., Felber R.,
RA   Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
RA   Chefdor F., Cooke R., Berger C., Monfort A., Casacuberta E.,
RA   Gibbons T., Weber N., Vandenbol M., Bargues M., Terol J., Torres A.,
RA   Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
RA   Heijnen L., Schwarz S., Scholler P., Heber S., Francs P., Bielke C.,
RA   Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
RA   Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
RA   Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
RA   Sehkon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
```

```
RA   Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
RA   Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
RA   Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA   Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
RA   Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
RA   Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshu C.,
RA   Antonoiu B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
RA   Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
RA   Swaby I.K., O'Shaugnessy A., Rodriguez M., Hoffman J., Till S.,
RA   Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
RA   Chen E., Marra M., Martienssen R., McCombie W.R.;
RT   "Sequence and analysis of chromosome 4 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 402:769-777(1999).
RN   [2]
RP   CONCEPTUAL TRANSLATION.
RA   Kieselbach T.;
RL   Unpublished observations (JUL-2001).
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- SIMILARITY: CONTAINS 5 KH DOMAINS.
CC   -!- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A GENE
CC       PREDICTION ERROR AND THE FUSION OF THE CDS FOR THIS PROTEIN WITH
CC       THE CDS FOR CHLOROPLAST PROTEASE HHOA.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL021710; CAA16717.1; ALT_SEQ.
DR   EMBL; AL161548; CAB78839.1; ALT_SEQ.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   Pfam; PF00013; KH-domain; 5.
DR   SMART; SM00322; KH; 5.
DR   PROSITE; PS50084; KH_TYPE_1; 5.
KW   Hypothetical protein; Nuclear protein; RNA-binding; Repeat.
FT   DOMAIN       35     99       KH 1.
FT   DOMAIN      138    210       KH 2.
FT   DOMAIN      311    380       KH 3.
FT   DOMAIN      394    455       KH 4.
FT   DOMAIN      535    599       KH 5.
SQ   SEQUENCE   606 AA;  65760 MW;  61F135BBB8647C0C CRC64;

  Query Match             7.1%;  Score 210.5;  DB 1;  Length 606;
  Best Local Similarity   19.6%;  Pred. No. 6.8e-06;
  Matches   89;  Conservative   86;  Mismatches  187;  Indels   93;  Gaps   14;

Qy    199 RLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAAEKSITIL---------- 247
          |:| |   || :||| |  |  |   |::|| |  : : | ::: |||
Db     39 RILCPIDVVGGVIGKSGKVINAIRHNTKAKIKVFDQLH-GCSQRVITIYCSVKEKQEEIG 97

Qy    248 ---STPEGTSAACKSILEIM---------HKEAQDIKFTEEIPLKILAHNNFVGRLIGKE 295
             |  |    |  ::|::          :    ::   :    ::|   :    ||||
Db     98 FTKSENEPLCCAQDALLKVYDAIVASDEENNTKTNVDRDDNKECRLLVPFSQSSSLIGKA 157

Qy    296 GRNLKKIEQDTDTKI-----TISPLQELTLYNPERTITVKGNVETCAK---AEEEIMKKI 347
          | |:|:| : |   :      :|    :     :  : : | |:  :   |   || ||
Db    158 GENIKRIRRRTRASVKVVSKDVSDPSHVCAMEYDNVVVISGEPESVKQALFAVSAIMYKI 217

Qy    348 --RESYENDIASMNL----------------------QAHL------IPGLNLNALGL 375
            ||:   |  | ::                      | |:      :|    ||| :
Db    218 NPRENIPLDSTSQDVPAASVIVPSDLSNSVYPQTGFYSNQDHILQQGAGVPSY-FNALSV 276

Qy    376 -----FPPTSGMPPPTSGPPSAMTPPYPQFEQSETETVHQFIPALSVGAIIGKQGQHIKQ 430
               :  |:  | |        :|  :    :||       |  ::  :||| |  ||:
Db    277 SDFQGYAETAANPVPVFASSLPVTHGFGGSSRSEELVFKVLCPLCNIMRVIGKGGSTIKR 336

Qy    431 LSRFAGASIKIAPAEAPDAKVR------MVIITGPPEAQFKAQGRIYGKIKEENFVSPKE 484
          :   :|: |     |  |:: :      ::|:|            :   :  : ::: ::

Db    337 IREASGSCI-----EVNDSRTKCGDDECVIIVTATESPDDMKSMAVEAVLLLQEYINDED 391

Qy    485 EVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVVVKITGHFY 544
             ::  : | |   | ||||  |  :||::  ::| : : : :  |     :|:::|
Db    392 AENVKMQLLVSSKVIGCVIGKSGSVINEIRKRTNANICISKGKKDD-----LVEVSGEV- 445

Qy    545 ACQVAQRKIQEILTQVKQHQQQKALQSGPPQSRRK 579
                  ::: | |:    ::  |      : ||
Db    446 ------SSVRDALIQIVLRLREDVLGDKDSVATRK 474


RESULT 7
PCB1_HUMAN
ID   PCB1_HUMAN     STANDARD;      PRT;   356 AA.
AC   Q15365; Q13157; Q14975;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Poly(rC)-binding protein 1 (Alpha-CP1) (hnRNP-E1) (Nucleic acid
DE   binding protein SUB2.3).
GN   PCBP1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95331278; PubMed=7607214;
RA   Leffers H., Dejgaard K., Celis J.E.;
RT   "Characterisation of two major cellular poly(rC)-binding human
RT   proteins, each containing three K-homologous (KH) domains.";
RL   Eur. J. Biochem. 230:447-453(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96016208; PubMed=7556077;
RA   Kiledjian M., Wang X., Liebhaber S.A.;
RT   "Identification of two KH domain proteins in the alpha-globin mRNP
RT   stability complex.";
RL   EMBO J. 14:4357-4364(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lymphocytes;
RX   MEDLINE=94203810; PubMed=8152927;
RA   Aasheim H.-C., Loukianova T., Deggerdal A., Smeland E.B.;
RT   "Tissue specific expression and cDNA structure of a human transcript
RT   encoding a nucleic acid binding [oligo(dC)] protein related to the
RT   pre-mRNA binding protein K.";
RL   Nucleic Acids Res. 22:959-964(1994).
CC   -!- FUNCTION: SINGLE-STRANDED NUCLEIC ACID BINDING PROTEIN THAT BINDS
CC       PREFERENTIALLY TO OLIGO DC.
CC   -!- SUBCELLULAR LOCATION: LOOSELY BOUND IN THE NUCLEUS. MAY SHUTTLE
CC       BETWEEN THE NUCLEUS AND THE CYTOPLASM.
CC   -!- TISSUE SPECIFICITY: ABUNDANTLY EXPRESSED IN SKELETAL MUSCLE,
CC       THYMUS AND PERIPHERAL BLOOD LEUCOCYTES WHILE A LOWER EXPRESSION IS
CC       OBSERVED IN PROSTATE, SPLEEN, TESTIS, OVARY, SMALL INTESTINE,
CC       HEART, LIVER, ADRENAL AND THYROID GLANDS.
CC   -!- PTM: PHOSPHORYLATED. THE NON-PHOSPHORYLATED FORM(S) EXHIBITED THE
CC       STRONGEST POLY(RC)-BINDING ACTIVITY.
CC   -!- SIMILARITY: CONTAINS 3 KH DOMAINS.
CC   -!- CAUTION: THERE IS PROBABLY A FRAMESHIFT ERROR IN THE NUCLEOTIDE
CC       SEQUENCE OF REF.3.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X78137; CAA55016.1; -.
DR   EMBL; U24223; AAA91317.1; -.
DR   EMBL; Z29505; CAA82631.1; ALT_FRAME.
```

```
DR   HSSP; Q07244; 1KHM.
DR   Genew; HGNC:8647; PCBP1.
DR   MIM; 601209; -.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1:
DR   Pfam; PF00013; KH-domain; 3.
DR   SMART; SM00322; KH; 3.
DR   PROSITE; PS50084; KH_TYPE_1; 3.
KW   Nuclear protein; RNA-binding; Ribonucleoprotein; DNA-binding;
KW   Phosphorylation; Repeat.
FT   DOMAIN       13     75       KH 1.
FT   DOMAIN       97    162       KH 2.
FT   DOMAIN      279    343       KH 3.
FT   CONFLICT    205    205       V -> A (IN REF. 2 AND 3).
SQ   SEQUENCE   356 AA;  37526 MW;  DC85477576DC5104 CRC64;

  Query Match             7.1%;  Score 209.5;  DB 1;  Length 356;
  Best Local Similarity   22.2%;  Pred. No. 3.8e-06;
  Matches   80;  Conservative   69;  Mismatches  110;  Indels  101;  Gaps   13;

Qy     196 LPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAAEKSITILSTPEGTSA 255
           | :|||:  : ||:|||||:| ::: | ::: ::|::    : |   : |  |: |  |:|
Db      14 LTIRLLMHGKEVGSIIGKKGESVKRIREESGARINI----SEGNCPERIITLTGP--TNA 67

Qy     256 ACKSILEIMHKEAQDIKFT---------EEIPLKILAHNNFVGRLIGKEGRNLKKIEQDT 306
             |:   |: |  :||  :           : |:::      | |||| |  :|:| : |
Db      68 IFKAFAMIIDKLEEDINSSMTNSTAASRPPVTLRLVVPATQCGSLIGKGGCKIKEIREST 127

Qy     307 DTKITISPLQELTLYNPERTITVKGNVETCAKAEEEI----MKKIRESYENDIASMNLQA 362
           :: ::   ::   : || ||: |  ::  :  ::|    :: : :| :  : ::  |
Db     128 GAQVQVA--GDMLPNSTERAITIAGVPQSVTECVKQICLVMLETLSQSPQGRVMTIPYQP 185

Qy     363 HLIPGLNLNALGLFPPTS--------------GMPPPT---SGPP-------------- 390
                     | :|               | |  |    |||
Db     186 ------------MPASSPVICAGGQDRCSDAVGYPHATHDLEGPPLDAYSIQGQHTISP 232

Qy     391 -------------------------SAMTPPYPQFE------QSETETVHQF-IPALSVG 418
                                    : :    |: :      : |:| |:  ||   :|
Db     233 LDLAKLNQVARQQSHFAMMHGGTGFAGIDSSSPEVKGYWASLDASTQTTHELTIPNNLIG 292

Qy     419 AIIGKQGQHIKQLSRFAGASIKIA-PAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEE 477
           |||:|| :| :: : :|| |||| | |      | | |||   :   ||  |  ::  |
Db     293 CIIGRQGANINEIRQMSGAQIKIANPVEGSSG--RQVTITGSAASISLAQYLINARLSSE 350


RESULT 8
PCB1_RABIT
ID   PCB1_RABIT     STANDARD;      PRT;   356 AA.
AC   O19048;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Poly(rC)-binding protein 1 (Alpha-CP1) (hnRNP-E1).
GN   PCBP1.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX   NCBI_TaxID=9986;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99207122; PubMed=10101190;
RA   Thiele B.J., Berger M., Huth A., Reimann I., Schwarz K., Thiele H.;
RT   "Tissue-specific translational regulation of alternative rabbit
RT   15-lipoxygenase mRNAs differing in their 3'-untranslated regions.";
RL   Nucleic Acids Res. 27:1828-1836(1999).
CC   -!- FUNCTION: SINGLE-STRANDED NUCLEIC ACID BINDING PROTEIN THAT BINDS
CC       PREFERENTIALLY TO OLIGO DC (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- SIMILARITY: CONTAINS 3 KH DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ003023; CAA05814.1; -.
DR   HSSP; Q07244; 1KHM.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   Pfam; PF00013; KH-domain; 3.
DR   SMART; SM00322; KH; 3.
DR   PROSITE; PS50084; KH_TYPE_1; 3.
KW   Nuclear protein; RNA-binding; Ribonucleoprotein; DNA-binding; Repeat.
FT   DOMAIN       13     75       KH 1.
FT   DOMAIN       97    162       KH 2.
FT   DOMAIN      279    343       KH 3.
SQ   SEQUENCE   356 AA;  37497 MW;  6D1A261276CA206D CRC64;

  Query Match             7.1%;  Score 209.5;  DB 1;  Length 356;
  Best Local Similarity   22.8%;  Pred. No. 3.8e-06;
  Matches   79;  Conservative   70;  Mismatches  123;  Indels   75;  Gaps   12;

Qy     196 LPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAAEKSITILSTPEGTSA 255
           | :|||:  : ||:|||||:| ::: | ::: ::|::    : |   : |  |: |  |:|
Db      14 LTIRLLMHGKEVGSIIGKKGESVKRIREESGARINI----SEGNCPERIITLTGP--TNA 67

Qy     256 ACKSILEIMHKEAQDIKFT---------EEIPLKILAHNNFVGRLIGKEGRNLKKIEQDT 306
             |:   |: |  :||  :           : |:::      | |||| |  :|:| : |
Db      68 IFKAFAMIIDKLEEDINSSMTNSTAASRPPVTLRLVVPATQCGSLIGKGGCKIKEIREST 127

Qy     307 DTKITISPLQELTLYNPERTITVKGNVETCAKAEEEI----MKKIRESYENDIASMNLQ- 361
           :: ::   ::   : || ||: |  ::  :  ::|    :: : :| :  : ::  |
Db     128 GAQVQVA--GDMLPNSTERAITIAGVPQSVTECVKQICLVMLETLSQSPQGRVMTIPYQP 185

Qy     362 -AHLIPGLNLNALGLFPPTSGMPPPT---SGPP--------------------------- 390
            | :                 :| |  |    |||
Db     186 MPASSPVICAGGQDRCSDAAGYPHATHDLEGPPLDAYSIQGQHTISPLDLAKLNQVARQQ 245

Qy     391 ------------SAMTPPYPQFE------QSETETVHQF-IPALSVGAIIGKQGQHIKQL 431
                       : :    |: :      : |:| |:  ||   :| |||:|| :| ::
Db     246 SHFAMMHGGTGFAGIDSSSPEVKGYWASLDASTQTTHELTIPNNLIGCIIGRQGANINEI 305

Qy     432 SRFAGASIKIA-PAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEE 477
           : :|| |||| | |      | | |||   :   ||  |  ::  |
Db     306 RQMSGAQIKIANPVEGSSG--RQVTITGSAASISLAQYLINARLSSE 350


RESULT 9
VGLN_HUMAN
ID   VGLN_HUMAN     STANDARD;      PRT;  1268 AA.
AC   Q00341; Q9UCY3;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Vigilin (High density lipoprotein-binding protein) (HDL-binding
DE   protein).
GN   HDLBP OR HBP OR VGL.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92291094; PubMed=1318310;
RA   McKnight G.L., Reasoner J., Gilbert T., Sundquist K.O., Hokland B.,
RA   McKernan P.A., Champagne J., Johnson C.J., Bailey M.C., Holly R.,
RA   O'Hara P.J., Oram J.F.;
RT   "Cloning and expression of a cellular high density lipoprotein-binding
RT   protein that is up-regulated by cholesterol loading of cells.";
RL   J. Biol. Chem. 267:12131-12141(1992).
RN   [2]
RP   SUBCELLULAR LOCATION.
RX   MEDLINE=96184515; PubMed=8605996;
```

```
RA   Kugler S., Grunweller A., Probst C., Klinger M., Muller P.K.,
RA   Kruse C.;
RT   "Vigilin contains a functional nuclear localisation sequence and is
RT   present in both the cytoplasm and the nucleus.";
RL   FEBS Lett. 382:330-334(1996).
CC   -!- FUNCTION: APPEARS TO PLAY A ROLE IN CELL STEROL METABOLISM.
CC       IT MAY FUNCTION TO PROTECT CELLS FROM OVER-ACCUMULATION OF
CC       CHOLESTEROL.
CC   -!- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.
CC   -!- SIMILARITY: CONTAINS 14 KH DOMAINS.
CC   -!- SIMILARITY: STRONG, TO C.ELEGANS C08H9.2.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M64098; AAA35962.1; -.
DR   Genew; HGNC:4857; HDLBP.
DR   MIM; 142695; -.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   Pfam; PF00013; KH-domain; 14.
DR   SMART; SM00322; KH; 14.
DR   PROSITE; PS50084; KH_TYPE_1; 14.
KW   HDL; Lipid transport; Cholesterol metabolism; RNA-binding; Repeat;
KW   Phosphorylation; Nuclear protein.
FT   DOMAIN      158    229       KH 1.
FT   DOMAIN      230    302       KH 2.
FT   DOMAIN      303    371       KH 3.
FT   DOMAIN      372    442       KH 4.
FT   DOMAIN      443    514       KH 5.
FT   DOMAIN      515    588       KH 6.
FT   DOMAIN      589    660       KH 7.
FT   DOMAIN      661    734       KH 8.
FT   DOMAIN      735    807       KH 9.
FT   DOMAIN      808    880       KH 10.
FT   DOMAIN      881    979       KH 11.
FT   DOMAIN      980   1059       KH 12.
FT   DOMAIN     1060   1134       KH 13.
FT   DOMAIN     1135   1209       KH 14.
FT   MOD_RES     295    295       PHOSPHORYLATION (POTENTIAL).
FT   MOD_RES     296    296       PHOSPHORYLATION (POTENTIAL).
SQ   SEQUENCE   1268 AA;  141439 MW;  374ECB83D13A7431 CRC64;

  Query Match             7.0%;  Score 207.5;  DB 1;  Length 1268;
  Best Local Similarity   23.9%;  Pred. No. 2.9e-05;
  Matches   95;  Conservative   69;  Mismatches  153;  Indels   81;  Gaps   18;

Qy    202 VPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAAEKSITILSTPEGTSAACKSIL 261
          :| :    :||| |  ::::  :| :|| : | ::       | |  | ||   |   :|
Db    157 IPKEHHRFVIGKNGEKLQDLELKTATKIQIPRPDD---PSNQIKITGTKEGIEKARHEVL 213

Qy    262 EIMHKEAQDIKFTEEIPLKILAHNNFVG---RLIGKEGRNLKKIEQDTDTKITISPLQEL 318
            |    || :  | : ::   |    |   ||:|       :| |:| |:| | |
Db    214 LI--SAEQDKRAVERLEVEKAFHPFIAGPYNRLVG-------EIMQETGTRINIPP---- 260

Qy    319 TLYNPERT-ITVKGNVETCAKAEEEIMKKIRESYENDIASMNL-----QAHLIPGLNLNA 372
             :  || |   |  |  |:|   | ||| |  :    :: :     |   : |   |:
Db    261 --PSVNRTEIVFTGEKEQLAQAVARI-KKIYEEKKKKTTTIAVEVKKSQHKYVIGPKGNS 317

Qy    373 LGLFPPTSG----MPPPTS--------GPP----SAMTPPYPQPEQSETETV------HQ 410
          |      :|    :||  |         | |      |:|  | :          :|     |:
Db    318 LQEILERTGVSVEIPPSDSISETVILRGEPEKLGQALTEVYAKANSFTVSSVAAPSWLHR 377

Qy    411 FIPALSVGAIIGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRI 470
          |          ||||:||:: ::::      : |   |  |     : : || |    || :|
Db    378 F--------IIGKKGQNLAKITQ-QMPKVHIEFTEGED----KITLEGPTEDVNVAQEQI 424

Qy    471 YGKIKE----ENFVSPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRD 526
```

```
          | :|:     ::|    : |   |:          ||| |  :| :::     | :|
Db    425 EGMVKDLINRMDYVEINIDHKFHRHL----------IGKSGANINRIKDQYKVSVRIP-- 472

Qy    527 QTPDENDQVVVKITGHFYACQVAQRKIQEILTQVKQHQ 564
            ||     :::| |      | |:|:: |: ::::  :
Db    473 --PDSEKSNLIRIEGDPQGVQQAKRELLELASRMENER 508


RESULT 10
PBP2_YEAST
ID   PBP2_YEAST     STANDARD;      PRT;   413 AA.
AC   P38151;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   PAB1-binding protein 2.
GN   PBP2 OR YBR233W OR YBR1531.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c;
RA   Dubois E., el Bakkoury M., Glansdorff N., Messenguy F., Pierard A.,
RA   Scherens B., Vierendeels F.;
RL   Submitted (AUG-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   PARTIAL CHARACTERIZATION.
RX   MEDLINE=99038243; PubMed=9819425;
RA   Mangus D.A., Amrani N., Jacobson A.;
RT   "Pbp1p, a factor interacting with Saccharomyces cerevisiae poly(A)-
RT   binding protein, regulates polyadenylation.";
RL   Mol. Cell. Biol. 18:7383-7396(1998).
CC   -!- SUBUNIT: INTERACTS WITH PAB1.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- SIMILARITY: CONTAINS 3 KH DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z36101; CAA85196.1; -.
DR   PIR; S46109; S46109.
DR   SGD; S0000437; PBP2.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   Pfam; PF00013; KH-domain; 3.
DR   SMART; SM00322; KH; 3.
DR   PROSITE; PS50084; KH_TYPE_1; 3.
KW   Nuclear protein; RNA-binding; Ribonucleoprotein; Repeat.
FT   DOMAIN       66    130       KH 1.
FT   DOMAIN      148    213       KH 2.
FT   DOMAIN      330    394       KH 3.
SQ   SEQUENCE   413 AA;  45782 MW;  638846509BCE1840 CRC64;

  Query Match             6.9%;  Score 204;  DB 1;  Length 413;
  Best Local Similarity   21.3%;  Pred. No. 1e-05;
  Matches   93;  Conservative   78;  Mismatches  164;  Indels  102;  Gaps   17;

Qy    155 PDEMAAQQNPLQQPRGR---------------RGLGQRGSSRQGSPGSVSKQKPCDLPL 198
          |  :    | |:: :|                     :       :   || :|    | |: |
Db     14 PTTVLVSPNTLKRKKGEDTSEEQLEAEIKRVALKDADSHSDNDHDSPDNV----PSDVHL 69

Qy    199 RLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAAEKSITILSTPEGTSAACK 258
          |:|   :    |:| :|||| |  :| ::|::       |  |: : :  | :  : |
Db     70 RMLCLVKHASLIVGHKGATISRIKSETSARINISNNIR-GVPERIVYVRGTCDDVAKAYG 128

Qy    259 SILEIMHKE------AQDIKFTEEIPLKILAHNNFVGRLIGKEGRNLKKIEQDTDTKITI 312
```

```
           |:  : :|       :||    || : :|  :: :| :||| |  |::||  :  |:
Db    129 MIVRALLEEHGNEDNGEDI----EISINLLIPHHLMGCIIGKRGSRLREIEDLSAAKLFA 184

Qy    313 SPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIASMNLQAHLIPGLNLNA 372
          || |   | : :| :|: | :    |   | : :          :| |    |  |:
Db    185 SPNQ--LLLSNDRILTINGVPDAIHIATFYISQTL----------LNFQME-SPQKNVKR 231

Qy    373 LGLFPPTSGMPPPTSGPPSAMTPPYPQF-----EQSETETV-----HQFIPALSVGAIIG 422
            : ||              .     ||      : |:  |:      ||: |  |   :
Db    232 SIYYQPT-----------------QFNSVLIDHSQPNTIFHQRNHQYHP--SDKLLSY 270

Qy    423 KQGQHIKQLS--------RFAGASIKIAPAEAPDAKV-RMVIITGPPEAQFKAQGRIYGK 473
          |  :::   |        ::  ||:  | |  |:  :  : :: ||  :        :
Db    271 KPNKNLPISSTLLSMATPQYTTASVANATAFQPNFVIPNVTVLDGPVISPAPGNHLLMNF 330

Qy    474 IKEENFVSPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDE-N 532
          :::| |:  |                | |||| || :| ::  :   :::   | | | :
Db    331 VQQEIFIDEK-------------FVGNVIGKDGKHINSVKESTGCSIII---QDPVEGS 373

Qy    533 DQVVVKITGHFYACQVA 549
            :  : | | | | | |
Db    374 SERRLTIRGTFMASQAA 390


RESULT 11
PCB2_HUMAN
ID   PCB2_HUMAN     STANDARD;      PRT;   365 AA.
AC   Q15366;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Poly(rC)-binding protein 2 (Alpha-CP2) (hnRNP-E2).
GN   PCBP2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95331278; PubMed=7607214;
RA   Leffers H., Dejgaard K., Celis J.E.;
RT   "Characterisation of two major cellular poly(rC)-binding human
RT   proteins, each containing three K-homologous (KH) domains.";
RL   Eur. J. Biochem. 230:447-453(1995).
CC   -!- FUNCTION: MAJOR CELLULAR POLY(RC)-BINDING PROTEIN. BINDS ALSO
CC       POLY(RU).
CC   -!- SUBCELLULAR LOCATION: LOOSELY BOUND IN THE NUCLEUS. MAY SHUTTLE
CC       BETWEEN THE NUCLEUS AND THE CYTOPLASM.
CC   -!- TISSUE SPECIFICITY: DETECTED IN ALL TISSUES EXAMINED.
CC   -!- PTM: PHOSPHORYLATED. THE NON-PHOSPHORYLATED FORM(S) EXHIBITED THE
CC       STRONGEST POLY(RC)-BINDING ACTIVITY.
CC   -!- SIMILARITY: CONTAINS 3 KH DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X78136; CAA55015.1; -.
DR   HSSP; Q07244; 1KHM.
DR   GeneW; HGNC:8648; PCBP2.
DR   MIM; 601210; -.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   Pfam; PF00013; KH-domain; 3.
DR   SMART; SM00322; KH; 3.
DR   PROSITE; PS50084; KH_TYPE_1; 3.
KW   Nuclear protein; RNA-binding; Ribonucleoprotein; DNA-binding;
KW   Phosphorylation; Repeat.
FT   DOMAIN       13     75       KH 1.
FT   DOMAIN       97    162       KH 2.
FT   DOMAIN      287    351       KH 3.
SQ   SEQUENCE   365 AA;  38580 MW;  43F035D76FDC2C63 CRC64;

  Query Match             6.9%;  Score 202.5;  DB 1;  Length 365;
  Best Local Similarity   24.0%;  Pred. No. 1.1e-05;
  Matches   87;  Conservative   58;  Mismatches  119;  Indels   99;  Gaps   13;

Qy    196 LPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAAEKSITILSTPEGTSA 255
          | :|||:  : ||:||||:| ::: : ::: ::|::    : |   : |  |: |  |:|
Db     14 LTIRLLMHGKEVGSIIGKKGESVKKMREESGARINI----SEGNCPERIITLAGP--TNA 67

Qy    256 ACKSILEIMHKEAQDIKFT---------EEIPLKILAHNNFVGRLIGKEGRNLKKIEQDT 306
            |:   |: |  :||  :           : |:::   :  | |||| |  :|:| : |
Db     68 IFKAFAMIIDKLEEDISSSMTNSTAASRPPVTLRLVVPASQCGSLIGKGGCKIKEIREST 127

Qy    307 DTKITISPLQELTLYNPERTITVKG---NVETCAKAEEEIMKKI---------------- 347
            :: ::   ::   : || ||: |   ::  | |    :| :
Db    128 GAQVQVA--GDMLPNSTERAITIAGIPQSIIECVKQICVVMLETLSQSPPKGVTIPYRPK 185

Qy    348 ------------RESYENDIASMNLQAHLIPGLNLNALGLFPPTSGMP----------- 383
                       | |  :| ||     |  | : ||     |   | |
Db    186 PSSSPVIFAGGQDRYSTGSDSASF---PHTTPSMCLN-----PDLEGPPLEAYTIQGQYA 237

Qy    384 ------------------PPTSG--------PPSAMTPPYPQFEQSETETVHQF-IPAL 415
                           | | |          |     |   : |   | |:  ||
Db    238 IPQPDLTKLHQLAMQQSHFPMTHGNTGFSGIESSSPEVKGYWGLDASAQTTSHELTIPND 297

Qy    416 SVGAIIGKQGQHIKQLSRFAGASIKIA-PAEAPDAKVRMVIITGPPEAQFKAQGRIYGKI 474
           :| |||:||  | :: : :|| |||| | |   :  | | |||   :   || |  ::
Db    298 LIGCIIGRQGAKINEIRQMSGAQIKIANPVEG--STDRQVTITGSAASISLAQYLINVRL 355

Qy    475 KEE 477
            |
Db    356 SSE 358


RESULT 12
PCB2_MOUSE
ID   PCB2_MOUSE     STANDARD;      PRT;   362 AA.
AC   Q61990; Q61383; Q62042;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Poly(rC)-binding protein 2 (Alpha-CP2) (Putative heterogeneous nuclear
DE   ribonucleoprotein X) (hnRNP X) (CTBP) (CBP).
GN   PCBP2 OR HNRNPX OR HNRPX OR CBP.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RX   MEDLINE=93376518; PubMed=8367306;
RA   Hahm K.B., Kim G., Turch C., Smale S.T.;
RT   "Isolation of a murine gene encoding a nucleic acid-binding protein
RT   with homology to hnRNP K.";
RL   Nucleic Acids Res. 21:3894-3894(1993).
RN   [2]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (ISOFORM 2).
RC   STRAIN=C57BL/6 X 129/Ola; TISSUE=Liver;
RX   MEDLINE=94268912; PubMed=8208614;
RA   Goller M., Funke B., Gehe-Becker C., Kroeger B., Lottspeich F.,
RA   Horak I.;
RT   "Murine protein which binds preferentially to oligo-C-rich single-
RT   stranded nucleic acids.";
RL   Nucleic Acids Res. 22:1885-1889(1994).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM 3).
RC   STRAIN=C57BL/6 X CBA; TISSUE=Liver;
RA   Horak I.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
RN   [4]
```

```
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RA   Makeyev A.V., Liebhaber S.A.;
RT   "Identification of two novel mammalian genes establishes a subfamily
RT   of KH-domain RNA-binding proteins.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: SINGLE-STRANDED NUCLEIC ACID BINDING PROTEIN THAT
CC       BINDS PREFERENTIALLY TO OLIGO DC.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1 (SHOWN HERE), 2 AND 3; ARE
CC       PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- SIMILARITY: CONTAINS 3 KH DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L19661; AAA03705.1; -.
DR   EMBL; X75947; CAA53546.1; -.
DR   EMBL; X97982; CAA66619.1; -.
DR   EMBL; AF236845; AAK14059.1; -.
DR   EMBL; AF236842; AAK14059.1; JOINED.
DR   EMBL; AF236843; AAK14059.1; JOINED.
DR   EMBL; AF236844; AAK14059.1; JOINED.
DR   HSSP; Q07244; 1KHM.
DR   MGD; MGI:108202; Pcbp2.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   Pfam; PF00013; KH-domain; 3.
DR   SMART; SM00322; KH; 3.
DR   PROSITE; PS50084; KH_TYPE_1; 3.
KW   Nuclear protein; RNA-binding; Ribonucleoprotein; DNA-binding;
KW   Repeat; Alternative splicing.
FT   DOMAIN       13     75       KH 1.
FT   DOMAIN       97     162      KH 2.
FT   DOMAIN      284     348      KH 3.
FT   VARSPLIC    194     224      MISSING (IN ISOFORM 2).
FT   VARSPLIC    263     275      MISSING (IN ISOFORM 3).
SQ   SEQUENCE   362 AA;  38221 MW;  70C8AF710E3BF3C0 CRC64;

  Query Match             6.8%;  Score 202;  DB 1;  Length 362;
  Best Local Similarity   24.2%;  Pred. No. 1.1e-05;
  Matches   87;  Conservative   60;  Mismatches  117;  Indels   96;  Gaps   14;

Qy    196 LPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAAEKSITILSTPEGTSA 255
          | :|||:  : ||:||||:| ::: : ::: ::|::    : |   : |  |: |  |:|
Db     14 LTIRLLMHGKEVGSIIGKKGESVKKMREESGARINI----SEGNCPERIITLAGP--TNA 67

Qy    256 ACKSILEIMHKEAQDIKFT---------EEIPLKILAHNNFVGRLIGKEGRNLKKIEQDT 306
            |:   |: |  :||  :           : |:::   :  | ||||  |  :|:| : |
Db     68 IFKAFAMIIDKLEEDISSSMTNSTAASRPPVTLRLVVPASQCGSLIGKGGCKIKEIREST 127

Qy    307 DTKITISPLQELTLYNPERTITVKG---NVETCAKAEEEIMKKI---------------- 347
          :: ::   ::   : || ||: |   ::  | |    :| :
Db    128 GAQVQVA--GDMLPNSTERAITIAGIPQSIIECVKQICVVMLESPPKGVTIPYRPKPSSS 185

Qy    348 ---------RESYENDIASMNLQAHLIPGLNLNALGLFPPTSGMP--------------- 383
                  | |  :| ||     |  | : ||     |   | |
Db    186 PVIFAGGQDRYSTGSDSASF---PHTTPSMCLN-----PDLEGPPLEAYTIQGQYAIPQP 237

Qy    384 ---------------PPTSGPP--SAMTPPYPQ-------FEQSETETVHQF-IPALSVG 418
                         | | |    | :    |:         : |   | |:  ||   :|
Db    238 DLTKLHQLAMQQSHFPMTHGNTGFSGIESSSPEVKGYWAGLDASAQTTSHELTIPNDLIG 297

Qy    419 AIIGKQGQHIKQLSRFAGASIKIA-PAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEE 477
           |||:||  | :: : :|| |||| | |   :  | | |||   :   || |  :: |
Db    298 CIIGRQGAKINEIRQMSGAQIKIANPVEG--STDRQVTITGSAASISLAQYLINVRLSSE 355


RESULT 13
```

```
VGLN_CHICK
ID   VGLN_CHICK     STANDARD;      PRT;  1270 AA.
AC   P81021;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Vigilin.
GN   HDLBP OR VGL.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryonic sternal cartilage;
RX   MEDLINE=92298988; PubMed=1606952;
RA   Schmidt C., Henkel B., Poeschl E., Zorbas H., Purschke W.G.,
RA   Gloe T.R., Mueller P.K.;
RT   "Complete cDNA sequence of chicken vigilin, a novel protein with
RT   amplified and evolutionary conserved domains.";
RL   Eur. J. Biochem. 206:625-634(1992).
CC   -!- SUBCELLULAR LOCATION: INTRACELLULAR.
CC   -!- SIMILARITY: CONTAINS 14 KH DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X65292; CAA46387.1; -.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   Pfam; PF00013; KH-domain; 14.
DR   SMART; SM00322; KH; 14.
DR   PROSITE; PS50084; KH_TYPE_1; 14.
KW   RNA-binding; Repeat.
FT   DOMAIN      150    188      KH 1.
FT   DOMAIN      219    260      KH 2.
FT   DOMAIN      291    333      KH 3.
FT   DOMAIN      360    402      KH 4.
FT   DOMAIN      431    473      KH 5.
FT   DOMAIN      504    545      KH 6.
FT   DOMAIN      577    619      KH 7.
FT   DOMAIN      651    693      KH 8.
FT   DOMAIN      724    766      KH 9.
FT   DOMAIN      798    840      KH 10.
FT   DOMAIN      872    913      KH 11.
FT   DOMAIN      970   1012      KH 12.
FT   DOMAIN     1051   1093      KH 13.
FT   DOMAIN     1126   1168      KH 14.
SQ   SEQUENCE   1270 AA;  142220 MW;  0A8844F91F9B8619 CRC64;

  Query Match             6.6%;  Score 194.5;  DB 1;  Length 1270;
  Best Local Similarity   21.8%;  Pred. No. 0.00019;
  Matches   97;  Conservative   85;  Mismatches  168;  Indels   95;  Gaps   17;

Qy    182 QGSPGSV--SKQKPCDLPLR--------LLVPTQFVGAIIGKEGATIRNITKQTQSKI-- 229
          :| |  |  :|::  :|  |         |::  :|   |||::|  || | ::    |
Db    483 EGDPQGVQQAKKELLELASRMENERTKDLIIEQKFHRTIIGQKGERIREIREKFPEVIIN 542

Qy    230 ---DVHRKENAGAAEKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNN 286
             |: :        |  |  |:     |   :: |  :  :  |:  :|:    | |
Db    543 FPDPAHKSD--------IVQLRGPKNEVEKCTKYMQKMVADLVENSFSISVPIFKQFHKN 594

Qy    287 FVGRLIGKEGRNLKKIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKK 346
             :||| | |:||| ::::||| :   :: |    :| :  :|  :|      |:
Db    595 ----IIGKGGANIKKIREESNTKIDLPGREQAT----QRQLLSQGREQTVKLLRHRIL-- 644

Qy    347 IRESYENDIASMNLQAHLIPGLNLNAL----GLFP--------------PTSG------- 381
```

```
Db    645 ---AIQKELANITEVEVSIPSKLHNSLIGAKGRFIRSIMEECGGVHIHFPTEGSGSATVT 701

Qy    382 --MPPPTSGPPSAMTPPYPQFEQSETETV--------HQFIPALSVGAIIGKQGQHIKQL 431

Db    702 IRAQPRTWRKPRSSCCTWAEEKQTKSYTVDLRAKPEYHKF--------LIGKGGGNIRKV 753

Qy    432 SRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKE-ENFVSPKEEVKLEA 490

Db    754 RDNTGARIIFPTSEDKDQE--LITIMGTEEAVKEAQKELEALIKNLDNVVEDSMVVDPKH 811

Qy    491 HIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVVVKITGHFYACQVAQ 550

Db    812 H--------RHFVIRRGQVLREIADEYGGVMVRLPTVSGTQSDKVTLK--GAKDCVEAAK 861

Qy    551 RKIQEILTQVKQHQQQKALQSGPPQ 575

Db    862 KRIQEIIEDL---EAQVTIECTIPQ 883
```

RESULT 14
ROK_HUMAN
ID ROK_HUMAN STANDARD; PRT; 463 AA.
AC Q07244; Q15671; Q96J62;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Heterogeneous nuclear ribonucleoprotein K (hnRNP K) (DC-stretch
DE binding protein) (CSBP) (Transformation upregulated nuclear protein)
DE (TUNP).
GN HNRPK OR HNRNPK.
OS Homo sapiens (Human), and
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606, 10116;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=Human;
RX MEDLINE=92107165; PubMed=1729596;
RA Matunis M.J., Michael W.M., Dreyfuss G.;
RT "Characterization and primary structure of the poly(C)-binding
RT heterogeneous nuclear ribonucleoprotein complex K protein.";
RL Mol. Cell. Biol. 12:164-171(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=Human;
RX MEDLINE=94149726; PubMed=8107114;
RA Dejgaard K., Leffers H., Rasmussen H.H., Madsen P., Kruse T.A.,
RA Gesser B., Nielsen H., Celis J.E.;
RT "Identification, molecular cloning, expression and chromosome mapping
RT of a family of transformation upregulated hnRNP-K proteins derived by
RT alternative splicing.";
RL J. Mol. Biol. 236:33-48(1994).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC SPECIES=Human; TISSUE=Muscle;
RA Strausberg R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES=Rat; STRAIN=Donryu; TISSUE=Liver;
RX MEDLINE=94173662; PubMed=8127654;
RA Ito K., Sato K., Endo H.;
RT "Cloning and characterization of a single-stranded DNA binding
RT protein that specifically recognizes deoxycytidine stretch.";
RL Nucleic Acids Res. 22:53-58(1994).
RN [5]
RP MASS SPECTROMETRY.
RC SPECIES=Human; TISSUE=Breast cancer;
RX MEDLINE=21829512; PubMed=11840567;
RA Harris R.A., Yang A., Stein R.C., Lucy K., Brusten L., Herath A.,
RA Parekh R., Waterfield M.D., O'Hare M.J., Neville M.A., Page M.J.,
RA Zvelebil M.J.;
RT "Cluster analysis of an extensive human breast cancer cell line
RT protein expression map database.";
RL Proteomics 2:212-223(2002).
RN [6]
RP STRUCTURE BY NMR OF 375-463.
RC SPECIES=Human;
RX MEDLINE=99299390; PubMed=10369774;
RA Baber J.L., Libutti D., Levens D., Tjandra N.;
RT "High precision solution structure of the C-terminal KH domain of
RT heterogeneous nuclear ribonucleoprotein K, a c-myc transcription
RT factor.";
RL J. Mol. Biol. 289:949-962(1999).
CC -!- FUNCTION: ONE OF THE MAJOR PRE-MRNA-BINDING PROTEINS. BINDS
CC TENACIOUSLY TO POLY(C) SEQUENCES. LIKELY TO PLAY A ROLE IN
CC THE NUCLEAR METABOLISM OF HNRNAS, PARTICULARLY FOR PRE-MRNAS THAT
CC CONTAIN CYTIDINE-RICH SEQUENCES. CAN ALSO BIND POLY(C) SINGLE-
CC STRANDED DNA.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC AND NUCLEAR; NUCLEOPLASM.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -!- PTM: PHOSPHORYLATED.
CC -!- MASS SPECTROMETRY: MW=50976.25; METHOD=MALDI.
CC -!- SIMILARITY: CONTAINS 3 KH DOMAINS.

DR EMBL; S74678; AAB20770.1; -.
DR EMBL; X72727; CAA51267.1; -.
DR EMBL; BC000355; AAH00355.1; -.
DR EMBL; BC014980; AAH14980.1; -.
DR EMBL; D17711; BAA04566.1; -.
DR PIR; A42058; A42058.
DR PIR; S41495; S41495.
DR PDB; 1KHM; 12-JAN-00.
DR TRANSFAC; T02853; -.
DR SWISS-2DPAGE; Q07244; HUMAN.
DR Genew; HGNC:5044; HNRPK.
DR MIM; 600712; -.
DR InterPro; IPR004087; KH_dom.
DR InterPro; IPR004088; KH_type_1.
DR Pfam; PF00013; KH-domain; 3.
DR SMART; SM00322; KH; 3.
DR PROSITE; PS50084; KH_TYPE_1; 3.
KW Nuclear protein; RNA-binding; Ribonucleoprotein; Repeat; DNA-binding;
KW Phosphorylation; Alternative splicing; 3D-structure.

FT	DOMAIN	42	92 KH 1.
FT	DOMAIN	144	197 KH 2.
FT	DOMAIN	236	273 RNA-BINDING (RGG-BOX).
FT	DOMAIN	387	439 KH 3.
FT	DOMAIN	54	421 2 X 22 AA APPROXIMATE REPEATS.
FT	REPEAT	54	76 1-1.
FT	REPEAT	399	421 1-2.
FT	DOMAIN	245	329 2 X 6 AA APPROXIMATE REPEATS.
FT	REPEAT	245	250 2-1.
FT	REPEAT	324	329 2-2.
FT	DOMAIN	59	407 5 X 4 AA REPEATS OF G-X-G-G.
FT	REPEAT	59	62 3-1.
FT	REPEAT	257	260 3-2.
FT	REPEAT	267	270 3-3.
FT	REPEAT	295	298 3-4.
FT	REPEAT	404	407 3-5.
FT	DOMAIN	289	294 POLY-PRO.
FT	DOMAIN	310	315 POLY-PRO.
FT	VARSPLIC	459	463 SGKFF -> ADVEGF (IN ISOFORM 2).
FT	CONFLICT	32	32 A -> D (IN REF. 2).

SQ SEQUENCE 463 AA; 50976 MW; 0F70EE169B2A064A CRC64;

```
  Query Match             6.4%;  Score 190;  DB 1;  Length 463;
  Best Local Similarity   20.7%;  Pred. No. 8.9e-05;
  Matches   87;  Conservative   54;  Mismatches  131;  Indels  148;  Gaps   13;

Qy    198 LRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAAEKSITILSTPEGTSAAC 257
          ||:|: ::  ||:||| |  |: :     : : |    ::   |: ::| :  |
Db     45 LRILLQSKNAGAVIGKGGKNIKALRTDYNASVSV---PDSSGPERILSISADIETIGEIL 101

Qy    258 KSILEIMHKEAQ--DIKFTEEIP--------------------LKILAHNNFVGRLIG 293
          | |:  : :  |      | ::|                    |::| | :  | :||
Db    102 KKIIPTLEEGLQLPSPTATSQLPLESDAVECLNYQHYKGSDFDCELRLLIHQSLAGGIIG 161

Qy    294 KEGRNLKKIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRES--- 350
          :|  :|:: ::| |  ||   ||   :: :| : : | :   :  : |:  | ||
Db    162 VKGAKIKELRENTQT--TIKLFQECCPHSTDRVVLIGGKPDRVVECIKIILDLISESPIK 219

Qy    351 ---------------YENDIASMNLQAHLIPGLNLNALGLF---PPTSG---MPP---- 384
                          |      : :    | :   | |   ||  |   |||
Db    220 GRAQPYDPNFYDETYDYGGFTMMFDDRRGRPVGFPMRGRGGFDRMPPGRGGRPMPPSRRD 279

Qy    385 -----PTSGPPS---------------AMTPPYPQ----------------------F 400
               |  |||                 : || |                      |
Db    280 YDDMSPRRGPPPPPPGRGGRGGSRARNLPLPPPPPPRGGDLMAYDRRGRPGDRYDGMVGF 339

Qy    401 EQSET-------------------------------------------ETVHQFIPALSV 417
            ||                                            |   ||
Db    340 SADETWDSAIDTWSPSEWQMAYEPQGGSGYDYSYAGGRGSYGDLGGPIITTQVTIPKDLA 399

Qy    418 GAIIGKQGQHIKQLSRFAGASIKI-APAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKE 476
          |:|||| || |||:   :||||||  | |   :: |:: |||  :    ||  :   :|:
Db    400 GSIIGKGGQRIKQIRHESGASIKIDEPLEG--SEDRIITITGTQDQIQNAQYLLQNSVKQ 457


RESULT 15
ROK_RABIT
ID   ROK_RABIT      STANDARD;      PRT;   463 AA.
AC   O19049;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Heterogeneous nuclear ribonucleoprotein K (hnRNP K).
GN   HNRPK.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX   NCBI_TaxID=9986;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99207122; PubMed=10101190;
RA   Thiele B.J., Berger M., Huth A., Reimann I., Schwarz K., Thiele H.;
RT   "Tissue-specific translational regulation of alternative rabbit
RT   15-lipoxygenase mRNAs differing in their 3'-untranslated regions.";
RL   Nucleic Acids Res. 27:1828-1836(1999).
CC   -!- FUNCTION: ONE OF THE MAJOR PRE-MRNA-BINDING PROTEINS. BINDS
CC       TENACIOUSLY TO POLY(C) SEQUENCES. LIKELY TO PLAY A ROLE IN THE
CC       NUCLEAR METABOLISM OF HNRNAS, PARTICULARLY FOR PRE-MRNAS THAT
CC       CONTAIN CYTIDINE-RICH SEQUENCES. CAN ALSO BIND POLY(C) SINGLE-
CC       STRANDED DNA.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC AND NUCLEAR; NUCLEOPLASM (BY
CC       SIMILARITY).
CC   -!- PTM: PHOSPHORYLATED (BY SIMILARITY).
CC   -!- SIMILARITY: CONTAINS 3 KH DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ003024; CAA05815.1; -.
DR   HSSP; Q07244; 1KHM.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   Pfam; PF00013; KH-domain; 3.
DR   SMART; SM00322; KH; 3.
DR   PROSITE; PS50084; KH_TYPE_1; 3.
KW   Nuclear protein; RNA-binding; Ribonucleoprotein; Repeat; DNA-binding;
KW   Phosphorylation.
FT   DOMAIN       42     92       KH 1.
FT   DOMAIN      144    197       KH 2.
FT   DOMAIN      236    273       RNA-BINDING (RGG-BOX).
FT   DOMAIN      387    439       KH 3.
FT   DOMAIN       54    421       2 X 22 AA APPROXIMATE REPEATS.
FT   REPEAT       54     76       1-1.
FT   REPEAT      399    421       1-2.
FT   DOMAIN      245    329       2 X 6 AA REPEATS OF D-R-R-G-R-P.
FT   REPEAT      245    250       2-1.
FT   REPEAT      324    329       2-2.
FT   DOMAIN       59    407       5 X 4 AA REPEATS OF G-X-G-G.
FT   REPEAT       59     62       3-1.
FT   REPEAT      257    260       3-2.
FT   REPEAT      267    270       3-3.
FT   REPEAT      295    298       3-4.
FT   REPEAT      404    407       3-5.
FT   DOMAIN      289    294       POLY-PRO.
FT   DOMAIN      308    315       POLY-PRO.
SQ   SEQUENCE   463 AA;  50960 MW;  A62A92E645EFE1B6 CRC64;

  Query Match             6.4%;  Score 190;  DB 1;  Length 463;
  Best Local Similarity   20.7%;  Pred. No. 8.9e-05;
  Matches   87;  Conservative   54;  Mismatches  131;  Indels  148;  Gaps   13;

Qy    198 LRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAAEKSITILSTPEGTSAAC 257
          ||:|: ::  ||:||| |  |: :     : : |    ::   |: ::| :  |
Db     45 LRILLQSKNAGAVIGKGGKNIKALRTDYNASVSV---PDSSGPERILSISADIETIGEIL 101

Qy    258 KSILEIMHKEAQ--DIKFTEEIP--------------------LKILAHNNFVGRLIG 293
          | |:  : :  |      | ::|                    |::| | :  | :||
Db    102 KKIIPTLEEGLQLPSPTATSQLPLESDAVECLNYQHFKGSDFDCELRLLIHQSLAGGIIG 161

Qy    294 KEGRNLKKIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRES--- 350
          :|  :|:: ::| |  ||   ||   :: :| : : | :   :  : |:  | ||
Db    162 VKGAKIKELRENTQT--TIKLFQECCPHSTDRVVLIGGKPDRVVECIKIILDLISESPIK 219

Qy    351 ---------------YENDIASMNLQAHLIPGLNLNALGLF---PPTSG---MPP---- 384
                          |      : :    | :   | |   ||  |   |||
Db    220 GRAQPYDPNFYDETYDYGGFTMMFDDRRGRPVGFPMRGRGGFDRMPPGRGGRPMPPSRRD 279

Qy    385 -----PTSGPPS---------------AMTPPYPQ----------------------F 400
               |  |||                 : || |                      |
Db    280 YDDMSPRRGPPPPPPGRGGRGGSRARNLPLPPPPPPRGGDLMAYDRRGRPGDRYDGMVGF 339

Qy    401 EQSET-------------------------------------------ETVHQFIPALSV 417
            ||                                            |   ||
Db    340 SADETWDSAIDTWSPSEWQMAYEPQGGSGYDYSYAGGRGSYGDLGGPIITTQVTIPKDLA 399

Qy    418 GAIIGKQGQHIKQLSRFAGASIKI-APAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKE 476
          |:|||| || |||:   :||||||  | |   :: |:: |||  :    ||  :   :|:
Db    400 GSIIGKGGQRIKQIRHESGASIKIDEPLEG--SEDRIITITGTQDQIQNAQYLLQNSVKQ 457


Search completed: April 24, 2003, 16:13:28
Job time : 16 secs
```

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: April 24, 2003, 16:11:07 ; Search time 40 Seconds
(without alignments)
1928.803 Million cell updates/sec

Title: US-09-897-778-176
Perfect score: 2956
Sequence: 1 MNKLYIGNLSENAAPSDLES.....VKQHQQQKALQSGPPQSRRK 579

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2956	100.0	579	21	AAB11328	Human lung cancer-
2	2956	100.0	579	23	ABB74960	Human lung tumour
3	2943	99.6	579	21	AAB11365	Human lung cancer
4	2943	99.6	579	23	ABB74997	Human lung tumour
5	2943	99.6	579	23	ABB75053	Human lung tumour
6	2943	99.6	579	23	ABB75054	Human lung tumour
7	2938	99.4	586	23	ABB75048	Human lung tumour
8	2814	95.2	619	22	ABG21963	Novel human diagno
9	2733	92.5	583	22	ABG12592	Novel human diagno
10	2190	74.1	577	20	AAY30649	A murine c-myc cod
11	1950.5	66.0	614	22	ABG06794	Novel human diagno
12	1946.5	65.8	620	22	AAU16163	Human novel secret
13	1868	63.2	594	22	ABG06795	Novel human diagno
14	1232	41.7	319	22	AAM93826	Human polypeptide,
15	949	32.1	261	22	AAU16161	Human novel secret
16	919	31.1	250	22	AAU16579	Human novel secret
17	835.5	28.3	558	22	ABB58367	Drosophila melanog
18	776.5	26.3	209	22	ABG21961	Novel human diagno
19	775.5	26.2	266	22	ABG12593	Novel human diagno
20	626	21.2	171	22	AAU16166	Human novel secret
21	551	18.6	148	22	ABG21962	Novel human diagno
22	527	17.8	192	22	AAU16164	Human novel secret
23	474	16.0	93	22	AAM78238	Human bone marrow
24	474	16.0	93	22	AAM38501	Peptide #12538 enc
25	402	13.6	171	22	AAU16583	Human novel secret
26	341	11.5	97	21	AAG03261	Human secreted pro
27	241	8.2	644	15	AAR58813	Human c-myc far up
28	231.5	7.8	313	21	AAG08900	Arabidopsis thalia
29	229.5	7.8	590	15	AAR58816	Human c-myc far up
30	229	7.7	386	22	ABB63279	Drosophila melanog
31	228.5	7.7	643	15	AAR58814	Human c-myc far up
32	227	7.7	45	22	ABB43427	Peptide #10933 enc
33	227	7.7	45	22	AAM64356	Human brain expres
34	227	7.7	45	22	AAM37315	Peptide #11352 enc
35	227	7.7	530	13	AAR28888	Ri paraneoplastic
36	226	7.6	1268	22	ABB61314	Drosophila melanog
37	226	7.6	1320	23	AAM47603	Drosophila cell cy
38	218	7.4	507	22	ABG23242	Novel human diagno
39	217.5	7.4	644	21	AAG39346	Arabidopsis thalia
40	213.5	7.2	403	21	AAB42547	Human ORFX ORF2311
41	213.5	7.2	403	22	AAB94825	Human protein sequ
42	207.5	7.0	1310	22	AAU32365	Novel human secret
43	204.5	6.9	781	21	AAG31213	Arabidopsis thalia
44	204.5	6.9	803	21	AAG31212	Arabidopsis thalia
45	204.5	6.9	846	21	AAG31211	Arabidopsis thalia

ALIGNMENTS

RESULT 1
AAB11328
ID AAB11328 standard; Protein; 579 AA.
XX
AC AAB11328;
XX
DT 21-FEB-2001 (first entry)
XX
DE Human lung cancer-associated protein L523S.
XX
KW Lung cancer; therapy; treatment; human; tumor; immunogenic; cytostatic;
KW vaccine; detection.
XX
OS Homo sapiens.
XX
PN WO200061612-A2.
XX
PD 19-OCT-2000.
XX
PF 03-APR-2000; 2000WO-US08896.
XX
PR 02-APR-1999; 99US-0285479.
PR 17-DEC-1999; 99US-0466396.
PR 30-DEC-1999; 99US-0476496.
PR 10-JAN-2000; 2000US-0480884.
PR 22-FEB-2000; 2000US-0510376.
XX
PA (CORI-) CORIXA CORP.
XX
PI Wang T, Fan L;
XX
DR WPI; 2000-628399/60.
DR N-PSDB; AAC65900.

```
XX
PT   Isolated polypeptide comprising an immunogenic portion of a lung tumor
PT   protein is used for detecting and monitoring progression of lung cancer
PT   in a patient -
XX
PS   Claim 3; Page 186-188; 261pp; English.
XX
CC   This invention describes a novel isolated polypeptide (I') which
CC   comprising an immunogenic portion of a lung tumor protein or variant (P2)
CC   which have cytostatic activity. The polypeptides and polynucleotides are
CC   used in compositions and vaccines to inhibit the development of cancer,
CC   especially lung cancer, in a patient. Methods described in the invention
CC   can be used to monitor the progression of a cancer by carrying out the
CC   detection at subsequent time points and comparing the results from the
CC   different time points. CD4+ and/or CD8+ T-Cells isolated from a patient
CC   are treated with P2, polynucleotides encoding P2 or antigen presenting
CC   cells expressing P2 and then administered to the patient to inhibit
CC   development of cancer.
XX
SQ   Sequence   579 AA;

  Query Match             100.0%;  Score 2956;  DB 21;  Length 579;
  Best Local Similarity   100.0%;  Pred. No. 9.7e-246;
  Matches  579;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGK 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGK 60

Qy     61 IELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 IELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETA 120

Qy    121 VVNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDEMAAQQNPLQQPRGRRGLGQRGSS 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 VVNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDEMAAQQNPLQQPRGRRGLGQRGSS 180

Qy    181 RQGSPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 RQGSPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAA 240

Qy    241 EKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEGRNLK 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 EKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEGRNLK 300

Qy    301 KIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIASMNL 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 KIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIASMNL 360

Qy    361 QAHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMTPPYPQFEQSETETVHQFIPALSVGAI 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 QAHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMTPPYPQFEQSETETVHQFIPALSVGAI 420

Qy    421 IGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEENFV 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 IGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEENFV 480

Qy    481 SPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVVVKIT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 SPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVVVKIT 540

Qy    541 GHFYACQVAQRKIQEILTQVKQHQQQKALQSGPPQSRRK 579
          |||||||||||||||||||||||||||||||||||||||
Db    541 GHFYACQVAQRKIQEILTQVKQHQQQKALQSGPPQSRRK 579


RESULT 2
ABB74960
ID   ABB74960 standard; Protein; 579 AA.
XX
AC   ABB74960;
XX
DT   01-MAY-2002  (first entry)
XX
DE   Human lung tumour L523S protein sequence SEQ ID NO:176.
XX
KW   Human; lung tumour; lung cancer; cytostatic; immunostimulant; vaccine;
KW   immune response.
XX
OS   Homo sapiens.
XX
PN   WO200200174-A2.
XX
PD   03-JAN-2002.
XX
PF   28-JUN-2001; 2001WO-US21065.
XX
PR   28-JUN-2000; 2000US-0606421.
PR   02-AUG-2000; 2000US-0630940.
PR   21-AUG-2000; 2000US-0643597.
PR   15-SEP-2000; 2000US-0662786.
PR   09-OCT-2000; 2000US-0685696.
PR   12-DEC-2000; 2000US-0735705.
PR   07-MAY-2001; 2001US-0850716.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Wang T,  Wang A,  Skeiky YAW,  Li SX,  Kalos MD,  Henderson RA;
PI   McNeill PD,  Fanger N,  Retter MW,  Marnerakis M,  Fanger GR;
PI   Vedvick TS,  Carter D,  Watanabe Y,  Peckham DW;
XX
DR   WPI; 2002-090513/12.
DR   N-PSDB; ABL49119.
XX
PT   Polynucleotides encoding lung tumor polypeptides, useful for treating
PT   lung cancer or stimulating an immune response -
XX
PS   Example 2; Page 267-268; 374pp; English.
XX
CC   The present invention describes human lung tumour proteins. Human lung
CC   tumour proteins and polynucleotides have cytostatic and immunostimulant
CC   activities, and can be used in vaccine production. Compositions
CC   comprising the lung tumour proteins, polynucleotides, antibodies,
CC   fusion proteins, T cell populations, or antigen presenting cells that
CC   express the lung tumour proteins are useful for treating lung cancer or
CC   stimulating an immune response. ABL48959 to ABL49300 and ABB74946 to
CC   ABB75070 represent sequences used in the exemplification of the present
CC   invention.
XX
SQ   Sequence   579 AA;

  Query Match             100.0%;  Score 2956;  DB 23;  Length 579;
  Best Local Similarity   100.0%;  Pred. No. 9.7e-246;
  Matches  579;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGK 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGK 60

Qy     61 IELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 IELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETA 120

Qy    121 VVNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDEMAAQQNPLQQPRGRRGLGQRGSS 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 VVNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDEMAAQQNPLQQPRGRRGLGQRGSS 180

Qy    181 RQGSPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 RQGSPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAA 240

Qy    241 EKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEGRNLK 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 EKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEGRNLK 300

Qy    301 KIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIASMNL 360
```

```
Db    301 KIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIASMNL 360

Qy    361 QAHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMTPPYPQFEQSETETVHQFIPALSVGAI 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 QAHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMTPPYPQFEQSETETVHQFIPALSVGAI 420

Qy    421 IGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEENFV 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 IGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEENFV 480

Qy    481 SPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVVVKIT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 SPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVVVKIT 540

Qy    541 GHFYACQVAQRKIQEILTQVKQHQQQKALQSGPPQSRRK 579
          |||||||||||||||||||||||||||||||||||||||
Db    541 GHFYACQVAQRKIQEILTQVKQHQQQKALQSGPPQSRRK 579


RESULT 3
AAB11365
ID   AAB11365 standard; Protein; 579 AA.
XX
AC   AAB11365;
XX
DT   21-FEB-2001  (first entry)
XX
DE   Human lung cancer associated antigen L523S.
XX
KW   Lung cancer; therapy; treatment; human; tumor; immunogenic; cytostatic;
KW   vaccine; detection.
XX
OS   Homo sapiens.
XX
PN   WO200061612-A2.
XX
PD   19-OCT-2000.
XX
PF   03-APR-2000; 2000WO-US08896.
XX
PR   02-APR-1999;   99US-0285479.
PR   17-DEC-1999;   99US-0466396.
PR   30-DEC-1999;   99US-0476496.
PR   10-JAN-2000; 2000US-0480884.
PR   22-FEB-2000; 2000US-0510376.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Wang T,  Fan L;
XX
DR   WPI; 2000-628399/60.
DR   N-PSDB; AAC66035.
XX
PT   Isolated polypeptide comprising an immunogenic portion of a lung tumor
PT   protein is used for detecting and monitoring progression of lung cancer
PT   in a patient -
XX
PS   Claim 3; Page 259-261; 261pp; English.
XX
CC   This invention describes a novel isolated polypeptide (I) which
CC   comprising an immunogenic portion of a lung tumor protein or variant (P2)
CC   which have cytostatic activity. The polypeptides and polynucleotides are
CC   used in compositions and vaccines to inhibit the development of cancer,
CC   especially lung cancer, in a patient. Methods described in the invention
CC   can be used to monitor the progression of a cancer by carrying out the
CC   detection at subsequent time points and comparing the results from the
CC   different time points. CD4+ and/or CD8+ T-Cells isolated from a patient
CC   are treated with P2, polynucleotides encoding P2 or antigen presenting
CC   cells expressing P2 and then administered to the patient to inhibit
CC   development of cancer.
XX
SQ   Sequence   579 AA;
```

```
  Query Match           99.6%;  Score 2943;  DB 21;  Length 579;
  Best Local Similarity  99.7%;  Pred. No. 1.3e-244;
  Matches  577;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGK 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGK 60

Qy     61 IELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 IELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETA 120

Qy    121 VVNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDEMAAQQNPLQQPRGRRGLGQRGSS 180
          |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db    121 VVNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDETAAQQNPLQQPRGRRGLGQRGSS 180

Qy    181 RQGSPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 RQGSPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAA 240

Qy    241 EKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEGRNLK 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 EKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEGRNLK 300

Qy    301 KIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIASMNL 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 KIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIASMNL 360

Qy    361 QAHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMTPPYPQFEQSETETVHQFIPALSVGAI 420
          |||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db    361 QAHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMTPPYPQFEQSETETVHLFIPALSVGAI 420

Qy    421 IGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEENFV 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 IGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEENFV 480

Qy    481 SPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVVVKIT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 SPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVVVKIT 540

Qy    541 GHFYACQVAQRKIQEILTQVKQHQQQKALQSGPPQSRRK 579
          |||||||||||||||||||||||||||||||||||||||
Db    541 GHFYACQVAQRKIQEILTQVKQHQQQKALQSGPPQSRRK 579


RESULT 4
ABB74997
ID   ABB74997 standard; Protein; 579 AA.
XX
AC   ABB74997;
XX
DT   01-MAY-2002  (first entry)
XX
DE   Human lung tumour L523S protein sequence SEQ ID NO:348.
XX
KW   Human; lung tumour; lung cancer; cytostatic; immunostimulant; vaccine;
KW   immune response.
XX
OS   Homo sapiens.
XX
PN   WO200200174-A2.
XX
PD   03-JAN-2002.
XX
PF   28-JUN-2001; 2001WO-US21065.
XX
PR   28-JUN-2000; 2000US-0606421.
PR   02-AUG-2000; 2000US-0630940.
PR   21-AUG-2000; 2000US-0643597.
PR   15-SEP-2000; 2000US-0662786.
PR   09-OCT-2000; 2000US-0685696.
PR   12-DEC-2000; 2000US-0735705.
```

PR 07-MAY-2001; 2001US-0850716.
XX
PA (CORI-) CORIXA CORP.
XX
PI Wang T, Wang A, Skeiky YAW, Li SX, Kalos MD, Henderson RA;
PI McNeill PD, Fanger N, Retter MW, Marnerakis M, Fanger GR;
PI Vedvick TS, Carter D, Watanabe Y, Peckham DW;
XX
DR WPI; 2002-090513/12.
DR N-PSDB; ABL49254.
XX
PT Polynucleotides encoding lung tumor polypeptides, useful for treating
PT lung cancer or stimulating an immune response -
XX
PS Example 2; Page 330-332; 374pp; English.
XX
CC The present invention describes human lung tumour proteins. Human lung
CC tumour proteins and polynucleotides have cytostatic and immunostimulant
CC activities, and can be used in vaccine production. Compositions
CC comprising the lung tumour proteins, polynucleotides, antibodies,
CC fusion proteins, T cell populations, or antigen presenting cells that
CC express the lung tumour proteins are useful for treating lung cancer or
CC stimulating an immune response. ABL48959 to ABL49300 and ABB74946 to
CC ABB75070 represent sequences used in the exemplification of the present
CC invention.
XX
SQ Sequence 579 AA;

Query Match 99.6%; Score 2943; DB 23; Length 579;
Best Local Similarity 99.7%; Pred. No. 1.3e-244;
Matches 577; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy      1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGK 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGK 60

Qy     61 IELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 IELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETA 120

Qy    121 VVNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDEMAAQQNPLQQPRGRRGLGQRGSS 180
          |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    121 VVNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDETAAQQNPLQQPRGRRGLGQRGSS 180

Qy    181 RQGSPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 RQGSPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAA 240

Qy    241 EKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEGRNLK 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 EKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEGRNLK 300

Qy    301 KIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIASMNL 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 KIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIASMNL 360

Qy    361 QAHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMTPPYPQFEQSETETVHQFIPALSVGAI 420
          |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    361 QAHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMTPPYPQFEQSETETVHLFIPALSVGAI 420

Qy    421 IGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEENFV 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 IGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEENFV 480

Qy    481 SPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVVVKIT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 SPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVVVKIT 540

Qy    541 GHFYACQVAQRKIQEILTQVKQHQQQKALQSGPPQSRRK 579
          |||||||||||||||||||||||||||||||||||||||
Db    541 GHFYACQVAQRKIQEILTQVKQHQQQKALQSGPPQSRRK 579
```

RESULT 5
ABB75053
ID ABB75053 standard; Protein; 579 AA.
XX
AC ABB75053;
XX
DT 01-MAY-2002 (first entry)
XX
DE Human lung tumour L523S recombinant protein sequence SEQ ID NO:446.
XX
KW Human; lung tumour; lung cancer; cytostatic; immunostimulant; vaccine;
KW immune response.
XX
OS Homo sapiens.
XX
PN WO200200174-A2.
XX
PD 03-JAN-2002.
XX
PF 28-JUN-2001; 2001WO-US21065.
XX
PR 28-JUN-2000; 2000US-0606421.
PR 02-AUG-2000; 2000US-0630940.
PR 21-AUG-2000; 2000US-0643597.
PR 15-SEP-2000; 2000US-0662786.
PR 09-OCT-2000; 2000US-0685696.
PR 12-DEC-2000; 2000US-0735705.
PR 07-MAY-2001; 2001US-0850716.
XX
PA (CORI-) CORIXA CORP.
XX
PI Wang T, Wang A, Skeiky YAW, Li SX, Kalos MD, Henderson RA;
PI McNeill PD, Fanger N, Retter MW, Marnerakis M, Fanger GR;
PI Vedvick TS, Carter D, Watanabe Y, Peckham DW;
XX
DR WPI; 2002-090513/12.
DR N-PSDB; ABL49297.
XX
PT Polynucleotides encoding lung tumor polypeptides, useful for treating
PT lung cancer or stimulating an immune response -
XX
PS Claim 2; Page 365-367; 374pp; English.
XX
CC The present invention describes human lung tumour proteins. Human lung
CC tumour proteins and polynucleotides have cytostatic and immunostimulant
CC activities, and can be used in vaccine production. Compositions
CC comprising the lung tumour proteins, polynucleotides, antibodies,
CC fusion proteins, T cell populations, or antigen presenting cells that
CC express the lung tumour proteins are useful for treating lung cancer or
CC stimulating an immune response. ABL48959 to ABL49300 and ABB74946 to
CC ABB75070 represent sequences used in the exemplification of the present
CC invention.
XX
SQ Sequence 579 AA;

Query Match 99.6%; Score 2943; DB 23; Length 579;
Best Local Similarity 99.7%; Pred. No. 1.3e-244;
Matches 577; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy      1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGK 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGK 60

Qy     61 IELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 IELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETA 120

Qy    121 VVNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDEMAAQQNPLQQPRGRRGLGQRGSS 180
          |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    121 VVNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDETAAQQNPLQQPRGRRGLGQRGSS 180

Qy    181 RQGSPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     181 RQGSPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAA 240

Qy     241 EKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEGRNLK 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 EKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEGRNLK 300

Qy     301 KIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIASMNL 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 KIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIASMNL 360

Qy     361 QAHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMTPPYPQFEQSETETVHQFIPALSVGAI 420
           |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db     361 QAHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMTPPYPQFEQSETETVHLFIPALSVGAI 420

Qy     421 IGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEENFV 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 IGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEENFV 480

Qy     481 SPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVVVKIT 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 SPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVVVKIT 540

Qy     541 GHFYACQVAQRKIQEILTQVKQHQQQKALQSGPPQSRRK 579
           |||||||||||||||||||||||||||||||||||||||
Db     541 GHFYACQVAQRKIQEILTQVKQHQQQKALQSGPPQSRRK 579
```

RESULT 6
ABB75054
ID ABB75054 standard; Protein; 579 AA.
XX
AC ABB75054;
XX
DT 01-MAY-2002 (first entry)
XX
DE Human lung tumour L523S recombinant protein sequence SEQ ID NO:449.
XX
KW Human; lung tumour; lung cancer; cytostatic; immunostimulant; vaccine;
KW immune response.
XX
OS Homo sapiens.
XX
PN WO200200174-A2.
XX
PD 03-JAN-2002.
XX
PF 28-JUN-2001; 2001WO-US21065.
XX
PR 28-JUN-2000; 2000US-0606421.
PR 02-AUG-2000; 2000US-0630940.
PR 21-AUG-2000; 2000US-0643597.
PR 15-SEP-2000; 2000US-0662786.
PR 09-OCT-2000; 2000US-0685696.
PR 12-DEC-2000; 2000US-0735705.
PR 07-MAY-2001; 2001US-0850716.
XX
PA (CORI-) CORIXA CORP.
XX
PI Wang T, Wang A, Skeiky YAW, Li SX, Kalos MD, Henderson RA;
PI McNeill PD, Fanger N, Retter MW, Marnerakis M, Fanger GR;
PI Vedvick TS, Carter D, Watanabe Y, Peckham DW;
XX
DR WPI; 2002-090513/12.
DR N-PSDB; ABL49299.
XX
PT Polynucleotides encoding lung tumor polypeptides, useful for treating
PT lung cancer or stimulating an immune response -
XX
PS Claim 2; Page 368-369; 374pp; English.
XX
CC The present invention describes human lung tumour proteins. Human lung
CC tumour proteins and polynucleotides have cytostatic and immunostimulant
CC activities, and can be used in vaccine production. Compositions
CC comprising the lung tumour proteins, polynucleotides, antibodies,
CC fusion proteins, T cell populations, or antigen presenting cells that
CC express the lung tumour proteins are useful for treating lung cancer or
CC stimulating an immune response. ABL48959 to ABL49300 and ABB74946 to
CC ABB75070 represent sequences used in the exemplification of the present
CC invention.
XX
SQ Sequence 579 AA;

Query Match 99.6%; Score 2943; DB 23; Length 579;
Best Local Similarity 99.7%; Pred. No. 1.3e-244;
Matches 577; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy       1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGK 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGK 60

Qy      61 IELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 IELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETA 120

Qy     121 VVNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDEMAAQQNPLQQPRGRRGLGQRGSS 180
           |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db     121 VVNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDETAAQQNPLQQPRGRRGLGQRGSS 180

Qy     181 RQGSPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAA 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 RQGSPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAA 240

Qy     241 EKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEGRNLK 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 EKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEGRNLK 300

Qy     301 KIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIASMNL 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 KIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIASMNL 360

Qy     361 QAHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMTPPYPQFEQSETETVHQFIPALSVGAI 420
           |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db     361 QAHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMTPPYPQFEQSETETVHLFIPALSVGAI 420

Qy     421 IGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEENFV 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 IGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEENFV 480

Qy     481 SPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVVVKIT 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 SPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVVVKIT 540

Qy     541 GHFYACQVAQRKIQEILTQVKQHQQQKALQSGPPQSRRK 579
           |||||||||||||||||||||||||||||||||||||||
Db     541 GHFYACQVAQRKIQEILTQVKQHQQQKALQSGPPQSRRK 579
```

RESULT 7
ABB75048
ID ABB75048 standard; Protein; 586 AA.
XX
AC ABB75048;
XX
DT 01-MAY-2002 (first entry)
XX
DE Human lung tumour L523S recombinant protein sequence SEQ ID NO:427.
XX
KW Human; lung tumour; lung cancer; cytostatic; immunostimulant; vaccine;
KW immune response.
XX
OS Homo sapiens.
XX
PN WO200200174-A2.
XX
PD 03-JAN-2002.
XX

PF 28-JUN-2001; 2001WO-US21065.
XX
PR 28-JUN-2000; 2000US-0606421.
PR 02-AUG-2000; 2000US-0630940.
PR 21-AUG-2000; 2000US-0643597.
PR 15-SEP-2000; 2000US-0662786.
PR 09-OCT-2000; 2000US-0685696.
PR 12-DEC-2000; 2000US-0735705.
PR 07-MAY-2001; 2001US-0850716.
XX
PA (CORI-) CORIXA CORP.
XX
PI Wang T, Wang A, Skeiky YAW, Li SX, Kalos MD, Henderson RA;
PI McNeill PD, Fanger N, Retter MW, Marnerakis M, Fanger GR;
PI Vedvick TS, Carter D, Watanabe Y, Peckham DW;
XX
DR WPI; 2002-090513/12.
DR N-PSDB; ABL49283.
XX
PT Polynucleotides encoding lung tumor polypeptides, useful for treating
PT lung cancer or stimulating an immune response -
XX
PS Claim 2; Page 354-355; 374pp; English.
XX
CC The present invention describes human lung tumour proteins. Human lung
CC tumour proteins and polynucleotides have cytostatic and immunostimulant
CC activities, and can be used in vaccine production. Compositions
CC comprising the lung tumour proteins, polynucleotides, antibodies,
CC fusion proteins, T cell populations, or antigen presenting cells that
CC express the lung tumour proteins are useful for treating lung cancer or
CC stimulating an immune response. ABL48959 to ABL49300 and ABB74946 to
CC ABB75070 represent sequences used in the exemplification of the present
CC invention.
XX
SQ Sequence 586 AA;

Query Match 99.4%; Score 2938; DB 23; Length 586;
Best Local Similarity 99.7%; Pred. No. 3.5e-244;
Matches 576; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 NKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGKI 61
Db 9 NKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGKI 68

Qy 62 ELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETAV 121
Db 69 ELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETAV 128

Qy 122 VNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDEMAAQQNPLQQPRGRRGLGQRGSSR 181
Db 129 VNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDETAAQQNPLQQPRGRRGLGQRGSSR 188

Qy 182 QGSPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAAE 241
Db 189 QGSPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAAE 248

Qy 242 KSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEGRNLKK 301
Db 249 KSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEGRNLKK 308

Qy 302 IEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIASMNLQ 361
Db 309 IEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIASMNLQ 368

Qy 362 AHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMTPPYPQFEQSETETVHQFIPALSVGAII 421
Db 369 AHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMTPPYPQFEQSETETVHLFIPALSVGAII 428

Qy 422 GKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEENFVS 481
Db 429 GKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEENFVS 488

Qy 482 PKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVVVKITG 541
Db 489 PKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVVVKITG 548

Qy 542 HFYACQVAQRKIQEILTQVKQHQQQKALQSGPPQSRRK 579
Db 549 HFYACQVAQRKIQEILTQVKQHQQQKALQSGPPQSRRK 586

RESULT 8
ABG21963
ID ABG21963 standard; Protein; 619 AA.
XX
AC ABG21963;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #21954.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder;
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS86150.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 52322; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 619 AA;

Query Match 95.2%; Score 2814; DB 22; Length 619;
Best Local Similarity 95.0%; Pred. No. 1.9e-233;
Matches 554; Conservative 8; Mismatches 17; Indels 4; Gaps 1;

```
Qy       1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGK 60
           |||||||||||||||||:|||||||||||||||||||||||||||||||||||:|:| |||
Db      37 MNKLYIGNLSENAAPSDLKSIFKDAKIPVSGPFLVKTGYAFVDCPDESWALRAIKAFSGK 96

Qy      61 IELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      97 IELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETA 156

Qy     121 VVNVTYSSKDQARQA----LDKLNGFQLENFTLKVAYIPDEMAAQQNPLQQPRGRRGLGQ 176
           |||||||||||||||    ||||||||||||||||||||| |||| ||||||||||
Db     157 VVNVTYSSKDQARQALDKLLDKLNGFQLENFTLKVAYIPDEMATQQNPFLQPRGRRGLGQ 216

Qy     177 RGSSRQGSPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKEN 236
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db     217 RGSSRQGSPGSVSKQKPCDSPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKEN 276

Qy     237 AGAAEKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEG 296
           |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     277 AGAAEKSITIPSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEG 336

Qy     297 RNLKKIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIA 356
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     337 RNLKKIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIA 396

Qy     357 SMNLQAHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMTPPYPQFEQSETETVHQFIPALS 416
           |||||||||||||||||||||||||||||||||||||||| || ||||:|||| ||||||
Db     397 SMNLQAHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMTPPSPQSEQSESETVHLFIPALS 456

Qy     417 VGAIIGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKE 476
           | |:| ||||||||||||||| |||| ||||||||||:| | |||:||||||||||||||
Db     457 VRALISKQGQHIKQLSRFAGASSKIAPVEAPDAKVRMVMIAGSPEARFKAQGRIYGKIKE 516

Qy     477 ENFVSPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVV 536
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     517 ENFVSPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVV 576

Qy     537 VKITGHFYACQVAQRKIQEILTQVKQHQQQKALQSGPPQSRRK 579
           ||||||||| ||||||| | ||||:|||||||||||||||||
Db     577 VKITGHFYACQVAQRKIXEFLTQVRQHQQQKALQSGPPQSRRK 619
```

RESULT 9
ABG12592
ID ABG12592 standard; Protein; 583 AA.
XX
AC ABG12592;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #12583.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS76779.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 42951; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 583 AA;

Query Match 92.5%; Score 2733; DB 22; Length 583;
Best Local Similarity 92.8%; Pred. No. 1.6e-226;
Matches 541; Conservative 6; Mismatches 32; Indels 4; Gaps 1;

```
Qy       1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGK 60
           ||| || |||||||| ||| | |||||||||||||| |||||||||| ||||||||||
Db       1 MNKPYIRNLSENAAPLDLEGILKDAKIPVSGPFLVKIGYAFVDCPDEIWALKAIEALSGK 60

Qy      61 IELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETA 120
           |||||||||||||||||||||||||||||||||||| ||||||||||:|||||||||
Db      61 IELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSXLVQYGVVESCQQVNTDSETA 120

Qy     121 VVNVTYSSKDQARQA----LDKLNGFQLENFTLKVAYIPDEMAAQQNPLQQPRGRRGLGQ 176
           ||||||||||||||    ||||||||||||||||||| |||| ||||||||||
Db     121 VVNVTYSSKDQARQALDKLLDKLNGFQLENFTLKVAYIPDEMATQQNPFLQPRGRRGLGQ 180

Qy     177 RGSSRQGSPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKEN 236
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db     181 RGSSRQGSPGSVSKQKPCDLPLCLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKEN 240

Qy     237 AGAAEKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEG 296
           |||||||||| ||||||||||||||||||||||||| || ||:|||||||| ||||||||||
Db     241 AGAAEKSITILSTPEGTSAACKSILEIMRKETQDVKFTEEIPLKILAQNNFVGRLIGKEG 300

Qy     297 RNLKKIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIA 356
           |||||||||||||||||||||  |||| ||| ||||||||||||||||||  ||||
Db     301 RNLKKIEQDTDTKITISPLQELMQYNPEHTITAKGNVETCAKAEEEIMKKIRERXENDIT 360

Qy     357 SMNLQAHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMTPPYPQFEQSETETVHQFIPALS 416
           |||||||||||||||||||||||||||||||||||||||| || ||||:|||| ||||||
Db     361 SMNLQAHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMTPPSPQSEQSESETVHLFIPALS 420

Qy     417 VGAIIGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKE 476
           | |:| ||||||||||||||| |||| ||||||||||:| | |||:||||||||||||||
Db     421 VRALISKQGQHIKQLSRFAGASSKIAPVEAPDAKVRMVMIAGSPEARFKAQGRIYGKIKE 480

Qy     477 ENFVSPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVV 536
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 ENFVSPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNFSSAEVVVPRDQTPDENDQVV 540

Qy     537 VKITGHFYACQVAQRKIQEILTQVKQHQQQKALQSGPPQSRRK 579
           |||||||||||||||||||||||||||||||||||||||||||
```

```
Db    541 VKITGHFYACQVAQRKIQEILTQVKQHQQQKALQSGPPQSRRK 583


RESULT 10
AAY30649
ID   AAY30649 standard; Protein; 577 AA.
XX
AC   AAY30649;
XX
DT   17-NOV-1999  (first entry)
XX
DE   A murine c-myc coding region determinant binding protein.
XX
KW   c-myc coding region determinant binding protein; CRD-BP; tumor;
KW   c-myc; endonucleolytic attack; half-life; breast cancer; colon cancer;
KW   pancreatic cancer.
XX
OS   Mus musculus.
XX
PN   WO9946594-A2.
XX
PD   16-SEP-1999.
XX
PF   05-MAR-1999;   99WO-US04897.
XX
PR   09-MAR-1998;   98US-0077372.
XX
PA   (WISC ) WISCONSIN ALUMNI RES FOUND.
XX
PI   Ross J;
XX
DR   WPI; 1999-551506/46.
DR   N-PSDB; AAZ10617.
XX
PT   Diagnosing presence or absence of a tumor in a human by examining c-myc
PT   coding region determinant-binding protein
XX
PS   Example; Fig 1A-D; 79pp; English.
XX
CC   The present sequence represents a murine c-myc coding region determinant
CC   binding protein (CRD-BP). The presence or absence of a tumor can be
CC   determined by determining the levels of CRD-BP present in the suspect
CC   tissue, where the CRD-BP shields c-myc RNA from endonucleolytic attack
CC   and so prolongs its half-life. The methods are used for diagnosing
CC   presence or absence of a tumor in a human, especially breast, colon
CC   and pancreatic cancer. They are also used to inhibit cancer cell
CC   growth.
XX
SQ   Sequence   577 AA;

  Query Match          74.1%;  Score 2190;  DB 20;  Length 577;
  Best Local Similarity  74.1%;  Pred. No. 1e-179;
  Matches  434;  Conservative   62;  Mismatches   74;  Indels   16;  Gaps    8;

Qy       1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGK 60
           |||||||||:|:  |:||| :| : ||  || ||||:|||||||||| ||:|||| |||
Db       1 MNKLYIGNLNESVTPADLEKVFAEHKISYSGQFLVKSGYAFVDCPDEHWAMKAIETFSGK 60

Qy      61 IELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETA 120
           :|| || :|:|||||||:|| ||:|||||| |:||||||| ||| ||:||||||:||||
Db      61 VELQGKRLEMEHSVPKKQRSRKIQIRNIPPQLRWEVLDSLLAQYGTVENCEQVNTESETA 120

Qy     121 VVNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDEMAAQQNPLQQPRGRR-GLGQRGS 179
           |||||||:::| |||: |||| |||| |||:|||||   |         ||| | | ||
Db     121 VVNVTYSNREQTRQAIMKLNGHQLENHALKVSYIPDEQITQ----GPENGRRGGFGSRGQ 176

Qy     180 SRQGSP---GSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKEN 236
           |||||    |: :||:| |:|||||||:||||||||||||||||||||||||||||||
Db     177 PRQGSPVAAGAPAKQQPVDIPLRLLVPTQYVGAIIGKEGATIRNITKQTQSKIDVHRKEN 236

Qy     237 AGAAEKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEG 296
           ||||||:|:: ||||| |:||| |||||||||:| |  :|:|||||||||||||||||||
Db     237 AGAAEKAISVHSTPEGCSSACKMILEIMHKEAKDTKTADEVPLKILAHNNFVGRLIGKEG 296

Qy     297 RNLKKIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIA 356
           |||||:||||:|||||| ||:|||||||||||||| :| | :||:|||||:||:||||:|
Db     297 RNLKKVEQDTETKITISSLQDLTLYNPERTITVKGAIENCCRAEQEIMKKVREAYENDVA 356

Qy     357 SMNLQAHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMT--PPYPQFEQS-ETETVHQFIP 413
           :|:||:|||||||| |:|||| :|   ||   |||::|   ||  | |: | | |  |||
Db     357 AMSLQSHLIPGLNLAAVGLFPASSSAVPP---PPSSVTGAAPYSSFMQAPEQEMVQVFIP 413

Qy     414 ALSVGAIIGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGK 473
           | :|||||||:|||||||||||| |||||| | ||:(||||:||||||||||||||||||
Db     414 AQAVGAIIGKKGQHIKQLSRFASASIKIAPPETPDSKVRMVVITGPPEAQFKAQGRIYGK 473

Qy     474 IKEENFVSPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDEND 533
           :|||||  |||||||| |||||: ||||||||||||||||||::|||||||||||||||
Db     474 LKEENFFGPKEEVKLETHIRVPASAAGRVIGKGGKTVNELQNLTAAEVVVPRDQTPDEND 533

Qy     534 QVVVKITGHFYACQVAQRKIQEILTQVKQHQQQKALQSGPPQSRRK 579
           ||:||| ||||| |:|||||::|| |||| | ||  ||   |:|||
Db     534 QVIVKIIGHFYASQMAQRKIRDILAQVKQ-QHQKG-QSNLAQARRK 577


RESULT 11
ABG06794
ID   ABG06794 standard; Protein; 614 AA.
XX
AC   ABG06794;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #6785.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS70981.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity
XX
PS   Claim 20; SEQ ID No 37153; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
```

```
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   614 AA;

  Query Match           66.0%;  Score 1950.5;  DB 22;  Length 614;
  Best Local Similarity 65.3%;  Pred. No. 4.8e-159;
  Matches  386;  Conservative   81;  Mismatches   95;  Indels   29;  Gaps    8;

Qy       1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGK 60
Db      17 MNKLYIGNLSPAVTADDLRQLFGDRKLPLAGQVLLKSGYAFVDYPDQNWAIRAIETLSGK 76

Qy      61 IELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETA 120
Db      77 VELHGKIMEVDYSVSKKLRSRKIQIRNIPPHLQWEVLDGLLAQYGTVENVEQVNTDTETA 136

Qy     121 VVNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDEMAAQQNPLQQPRGRRGLGQRGSS 180
Db     137 VVNVTYATREEAKIAMEKLSGHQFENYSFKISYIPDEEVSSPSPPQ--RAQR--GDHSSR 192

Qy     181 RQG-SPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGA 239
Db     193 EQGHAPGGTSQARQIDFPLRILVPTQFVGAIIGKEGLTIKNITKQTQSRVDIHRKENSGA 252

Qy     240 AEKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEGRNL 299
Db     253 AEKPVTIHATPEGTSEACRMILEIMQKEADETKLAEEIPLKILAHNGLVGRLIGKEGRNL 312

Qy     300 KKIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIASMN 359
Db     313 KKIEHETGTKITISSLQDLSIYNPERTITVKGTVEACASAEIEIMKKLREAFENDMLAVN 372

Qy     360 LQAHLIPGDNLNALGLF--------PPTS--GMPPPT--------SGPPSAMTP-----P 396
Db     373 QQANLIPGDNLSALGIFSTGLSVLSPPAGPRGAPPAAPYHPFTTHSGYFSSLYPHHQFGP 432

Qy     397 YPQFEQ-SETETVHQFIPALSVGAIIGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVI 455
Db     433 FPHHHSYPEQEIVNLFIPTQAVGAIIGKKGAHIKQLARFAGASIKIAPAEGPDVSERMVI 492

Qy     456 ITGPPEAQFKAQGRIYGKIKEENFVSPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQN 515
Db     493 ITGPPEAQFKAQGRIFGKLKEENFFNPKEEVKLEAHIRVPSSTAGRVIGKGGKTVNELQN 552

Qy     516 LSSAEVVVPRDQTPDENDQVVVKITGHFYACQVAQRKIQEILTQVKQHQQQ 566
Db     553 LTSAEVIVPRDQTPDENEEVIVRIIGHFFASQTAQRKIREIVQQVKQQEQK 603


RESULT 12
AAU16163
ID   AAU16163 standard; Protein; 620 AA.
XX
AC   AAU16163;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human novel secreted protein, Seq ID 1116.
XX
KW   Human; immunosuppressive; antiarthritic; antirheumatic;
KW   cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;
KW   neuroprotective; antibacterial; virucide; fungicide; opthalmalogical;
KW   vulnerary; secreted protein; rheumatoid arthritis;
KW   hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
KW   cerebrovascular disorder; cerebral ischaemia; angiogenesis;
KW   nervous system disorder; Alzheimer's disease; infection; ocular disorder;
KW   corneal infection; wound healing; epithelial cell proliferation;
KW   skin ageing; food additive; preservative; antiproliferative.
XX
OS   Homo sapiens.
XX
PN   WO200155322-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01341.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
```

PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-488783/53.
DR N-PSDB; AAS26150.
XX
PT New nucleic acid molecules encoding 461 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives -
XX
PS Claim 11; SEQ ID No 1116; 980pp; English.
XX
CC The invention relates to isolated nucleic acid molecules and their
CC encoded secreted proteins. The nucleic acids and proteins are used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They
CC are also used in diagnosing a pathological condition or susceptibility
CC to a pathological condition. Antibodies to the proteins can also
CC be used in alleviating symptoms associated with the disorders and in
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
CC immunosorbant assays (ELISA). Disorders which are diagnosed or treated
CC include autoimmune diseases e.g. rheumatoid arthritis,
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
CC and ocular disorders e.g. corneal infection, and many other
CC disorders listed in the specification. The polypeptides can also
CC be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, cofactors and other nutritional components. The present
CC sequence represents a novel secreted protein of the invention.

Query Match 65.8%; Score 1946.5; DB 22; Length 620;
Best Local Similarity 65.1%; Pred. No. 1.1e-158;
Matches 385; Conservative 81; Mismatches 96; Indels 29; Gaps 8;

```
Qy       1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGK 60
           |||||||||||      ||  :| | |:|::|  |:|:|||||| ||::||::||| ||||
Db      23 MNKLYIGNLSPAVTADDLRQLFGDRKLPLAGQVLLKSGYAFVDYPDQNWAIRAIETLSGK 82

Qy      61 IELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETA 120
           :|||||:||::|| |: | ||:|||||||||||||||| || ||| ||: ||||||:|||
Db      83 VELHGKIMEVDYSVSKKLRSRKIQIRNIPPHLQWEVLDGLLAQYGTVENVEQVNTDTETA 142

Qy     121 VVNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDEMAAQQNPLQQPRGRRGLGQRGSS 180
           ||||||:::::|: |::||:| | ||:: |::|||||  :  :| |  | :|  |  |
Db     143 VVNVTYATREEAKIAMEKLSGHQFENYSFKISYIPDEEVSSPSPPQ--RAQR--GDHSSR 198

Qy     181 RQG-SPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGA 239
            || :||  |: :  | |||:|||||||||||||||| ||:||||||||::|:|||||:||
Db     199 EQGHAPGGTSQARQIDFPLRILVPTQFVGAIIGKEGLTIKNITKQTQSRVDIHRKENSGA 258

Qy     240 AEKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEGRNL 299
           ||| :|| :|||||| ||: ||||| ||| : |  |||||||||||  ||||||||||||
Db     259 AEKPVTIHATPEGTSEACRMILEIMQKEADETKLAEEIPLKILAHNGLVGRLIGKEGRNL 318

Qy     300 KKIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIASMN 359
           |||| :| |||||| ||:|::|||||||||||| || || || || |||||:||::||| ::|
Db     319 KKIEHETGTKITISSLQDLSIYNPERTITVKGTVEACASAEIEIMKKLREAFENDMLAVN 378

Qy     360 LQAHLIPGLNLNALGLF--------PPTS--GMPPPT--------SGPPSAMTP-----P 396
           ||:|||||||:|||:|        ||    | ||          ||  |:: |    . |
Db     379 QQANLIPGLNLSALGIFSTGLSVLSPPAGPRGAPPAAPYHPFTTHSGYFSSLYPHHQFGP 438

Qy     397 YPQFEQ-SETETVHQFIPALSVGAIIGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVI 455
```

```
Db    439 FPHHHSYPEQEIVNLFIPTQAVGAIIGKKGAHIKQLARFAGASIKIAPAEGPDVSERMVI 498

Qy    456 ITGPPEAQFKAQGRIYGKIKEENFVSPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQN 515

Db    499 ITGPPEAQFKAQGRIFGKLKEENFFNPKEEVKLEXHIRVPSSTAGRVIGKGGKTVNELQN 558

Qy    516 LSSAEVVVPRDQTPDENDQVVVKITGHFYACQVAQRKIQEILTQVKQHQQQ 566

Db    559 LTSAEVIVPRDQTPDENEEVIVRIIGHFFASQTAQRKIREIVQQVKQQEQK 609


RESULT 13
ABG06795
ID   ABG06795 standard; Protein; 594 AA.
XX
AC   ABG06795;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #6786.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS70982.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity
XX
PS   Claim 20; SEQ ID No 37154; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   594 AA;

  Query Match           63.2%;  Score 1868;  DB 22;  Length 594;
  Best Local Similarity   64.8%;  Pred. No. 5.9e-152;
  Matches  368;  Conservative   78;  Mismatches   96;  Indels   26;  Gaps    6;

Qy      1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGK 60

Db     40 MNKLYIGNLSPAVTADDLRQLFGDRKLPLAGQVLLKSGYAFVDYPDQNWAIRAIETLSGK 99

Qy     61 IELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETA 120

Db    100 VELHGKIMEVDYSVSKKLRSRKIQIRNIPPHLQWEVLDGLLAQYGTVENVEQVNTDTETA 159

Qy    121 VVNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDEMAAQQNPLQQPRGRRGLGQRGSS 180

Db    160 VVNVTYATREEAKIAMEKLSGHQFENYSFKISYIPDEEVSSPSPPQ--RAQR--GDHSSR 215

Qy    181 RQG-SPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGA 239

Db    216 EQGHAPGGTSQARQIDFPLRILVPTQFVGAIIGKEGLTIKNITKQTQSRVDIHRKENSGA 275

Qy    240 AEKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEGRNL 299

Db    276 AEKPVTIHATPEGTSEACRMILEIMQKEADETKLAEEIPLKILAHNGLVGRLIGKEGRNL 335

Qy    300 KKIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIASMN 359

Db    336 KKIEHETGTKITISSLQDLSIYNPERTITVKGTVEACASAEIEIMKKLREAFENDMLAVN 395

Qy    360 LQAHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMTPPYPQFEQ-SETETVHQFIPALSVG 418

Db    396 THS----------GYF---------SSLYPHHQFGPFPHHHSYPEQEIVNLFIPTQAVG 435

Qy    419 AIIGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEEN 478

Db    436 AIIGKKGAHIKQLARFAGASIKIAPAEGPDVSERMVIITGPPEAQFKAQGRIFGKLKEEN 495

Qy    479 FVSPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVVVK 538

Db    496 FFNPKEEVKLEAHIRVPSSTAGRVIGKGGKTVNELQNLTSAEVIVPRDQTPDENEEVIVR 555

Qy    539 ITGHFYACQVAQRKIQEILTQVKQHQQQ 566

Db    556 IIGHFFASQTAQRKIREIVQQVKQQEQK 583


RESULT 14
AAM93826
ID   AAM93826 standard; Protein; 319 AA.
XX
AC   AAM93826;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human polypeptide, SEQ ID NO: 3887.
XX
KW   Human; full length cDNA; cDNA synthesis; oligo-capping.
XX
OS   Homo sapiens.
XX
PN   EP1130094-A2.
XX
PD   05-SEP-2001.
XX
PF   07-JUL-2000; 2000EP-0114089.
XX
PR   08-JUL-1999;   99JP-0194486.
PR   11-JAN-2000; 2000JP-0118774.
PR   02-MAY-2000; 2000JP-0183765.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Nishikawa T,  Isogai T,  Hayashi K,  Ishii S,  Kawai Y;
```

```
PI   Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
DR   WPI; 2001-524255/58.
DR   N-PSDB; AAK94782.
XX
PT   830 Primers useful for synthesizing full length cDNA clones and their
PT   use in genetic manipulation -
XX
PS   Claim 8; SEQ ID NO 3887; 1380pp + sequence listing; English.
XX
CC   The invention relates to primers for synthesising full length cDNA
CC   clones. 830 cDNA molecules encoding a human protein have been
CC   isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC   molecules have been determined. Primers for synthesising the full length
CC   cDNA are useful for clarifying the function of the protein encoded by
CC   the cDNA. The full length clones were obtained by construction of full
CC   length enriched cDNA libraries that were synthesised by the oligo-capping
CC   method. The primers enable the production of the full length cDNA easily
CC   without any special methods. The present sequence is a polypeptide
CC   encoded by a full length human cDNA of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in CD-ROM format directly from EPO.
XX
SQ   Sequence   319 AA;

  Query Match          41.7%;  Score 1232;  DB 22;  Length 319;
  Best Local Similarity   76.5%;  Pred. No. 1.5e-97;
  Matches  247;  Conservative   32;  Mismatches   36;  Indels    8;  Gaps    5;

Qy     260 ILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEGRNLKKIEQDTDTKITISPLQELT 319
           |||||||||:| |  :|:|||||||||||||||||||||||:||||:|||||| ||:||
Db       2 ILEIMHKEAKDTKTADEVPLKILAHNNFVGRLIGKEGRNLKKVEQDTETKITISSLQDLT 61

Qy     320 LYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIASMNLQAHLIPGLNLNALGLFPPT 379
           ||||||||||| :| | :||:|||||:||:||||:|:|:||:|||||||| |:|||| :
Db      62 LYNPERTITVKGAIENCCRAEQEIMKKVREAYENDVAAMSLQSHLIPGLNLAAVGLFPAS 121

Qy     380 SGMPPPTSGPPSAMT--PPYPQFEQS-ETETVHQFIPALSVGAIIGKQGQHIKQLSRFAG 436
           |   ||   |||::|   ||  | |: | | |  |||| :|||||||:||||||||||
Db     122 SSAVPP---PPSSVTGAAPYSSFMQAPEQEMVQVFIPAQAVGAIIGKKGQHIKQLSRFAS 178

Qy     437 ASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEENFVSPKEEVKLEAHIRVPS 496
           ||||||| |||:|||||||||||||||||||||||:|||||  ||||||| |||||:
Db     179 ASIKIAPPETPDSKVRMVIITGPPEAQFKAQGRIYGKLKEENFFGPKEEVKLETHIRVPA 238

Qy     497 FAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVVVKITGHFYACQVAQRKIQEI 556
           |||||||||||||||||::|||||||||||||||:||| ||||| |:|||||::|
Db     239 SAAGRVIGKGGKTVNELQNLTAAEVVVPRDQTPDENDQVIVKIIGHFYASQMAQRKIRDI 298

Qy     557 LTQVKQHQQQKALQSGPPQSRRK 579
           | |||| | ||  ||   |:|||
Db     299 LAQVKQ-QHQKG-QSNQAQARRK 319


RESULT 15
AAU16161
ID   AAU16161 standard; Protein; 261 AA.
XX
AC   AAU16161;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human novel secreted protein, Seq ID 1114.
XX
KW   Human; immunosuppressive; antiarthritic; antirheumatic;
KW   cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;
KW   neuroprotective; antibacterial; virucide; fungicide; opthalmalogical;
KW   vulnerary; secreted protein; rheumatoid arthritis;
KW   hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
KW   cerebrovascular disorder; cerebral ischaemia; angiogenesis;
KW   nervous system disorder; Alzheimer's disease; infection; ocular disorder;
KW   corneal infection; wound healing; epithelial cell proliferation;
KW   skin ageing; food additive; preservative; antiproliferative.
XX
OS   Homo sapiens.
XX
PN   WO200155322-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01341.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
```

PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX

PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-488783/53.
DR N-PSDB; AAS26148.
XX
PT New nucleic acid molecules encoding 461 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives -
XX
PS Claim 11; SEQ ID No 1114; 980pp; English.
XX
CC The invention relates to isolated nucleic acid molecules and their
CC encoded secreted proteins. The nucleic acids and proteins are used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They
CC are also used in diagnosing a pathological condition or susceptibility
CC to a pathological condition. Antibodies to the proteins can also
CC be used in alleviating symptoms associated with the disorders and in
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
CC immunosorbant assays (ELISA). Disorders which are diagnosed or treated
CC include autoimmune diseases e.g. rheumatoid arthritis,
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
CC and ocular disorders e.g. corneal infection, and many other
CC disorders listed in the specification. The polypeptides can also
CC be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, cofactors and other nutritional components. The present
CC sequence represents a novel secreted protein of the invention.

Query Match 32.1%; Score 949; DB 22; Length 261;
Best Local Similarity 74.5%; Pred. No. 2.6e-73;
Matches 193; Conservative 26; Mismatches 32; Indels 8; Gaps 5;

```
Qy    324 ERTITVKGNVETCAKAEEEIMKKIRESYENDIASMNLQAHLIPGLNLNALGLFPPTSGMP 383
          |||||||| :| | :||:|||||:||:||||:|:|:||:|||||||| |:|||| :|
Db      8 ERTITVKGAIENCCRAEQEIMKKVREAYENDVAAMSLQSHLIPGLNLAAVGLFPASSSAV 67

Qy    384 PPTSGPPSAMT--PPYPQFEQS-ETETVHQFIPALSVGAIIGKQGQHIKQLSRFAGASIK 440
          ||   |||::|   ||  | |: | | |  |||| :|||||||:|||||||||| ||||
Db     68 PP---PPSSVTGAAPYSSFMQAPEQEMVQVFIPAQAVGAIIGKKGQHIKQLSRFASASIK 124

Qy    441 IAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEENFVSPKEEVKLEAHIRVPSFAAG 500
          ||| | ||:||||||||||||||||||||||:|||||  |||||||| |||||: |||
Db    125 IAPPETPDSKVRMVIITGPPEAQFKAQGRIYGKLKEENFFGPKEEVKLETHIRVPASAAG 184

Qy    501 RVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVVVKITGHFYACQVAQRKIQEILTQV 560
          |||||||||||||||::|||||||||||||||||:||| ||||| |:||||::|| ||
Db    185 RVIGKGGKTVNELQNLTAAEVVVPRDQTPDENDQVIVKIIGHFYASQMAQRKIRDILAQV 244

Qy    561 KQHQQQKALQSGPPQSRRK 579
          || | ||  ||   |:|||
Db    245 KQ-QHQKG-QSNQAQARRK 261
```

Search completed: April 24, 2003, 16:15:28
Job time : 43 secs

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: April 24, 2003, 16:11:07 ; Search time 22 Seconds
(without alignments)
2108.871 Million cell updates/sec

Title: US-09-897-778-176
Perfect score: 2956
Sequence: 1 MNKLYIGNLSENAAPSDLES.....VKQHQQQKALQSGPPQSRRK 579

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 301932 seqs, 80129803 residues

Total number of hits satisfying chosen parameters: 301932

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2956	100.0	579	10	US-09-735-705-176	Sequence 176, App
2	2956	100.0	579	10	US-09-850-716A-176	Sequence 176, App
3	2956	100.0	579	10	US-09-897-778-176	Sequence 176, App
4	2943	99.6	579	10	US-09-735-705-348	Sequence 348, App
5	2943	99.6	579	10	US-09-850-716A-348	Sequence 348, App
6	2943	99.6	579	10	US-09-897-778-348	Sequence 348, App
7	2943	99.6	579	10	US-09-897-778-446	Sequence 446, App
8	2943	99.6	579	10	US-09-897-778-449	Sequence 449, App
9	2938	99.4	586	10	US-09-850-716A-427	Sequence 427, App
10	2938	99.4	586	10	US-09-897-778-427	Sequence 427, App
11	2190	74.1	577	10	US-09-873-637-2	Sequence 2, Appli
12	1946.5	65.8	620	10	US-09-764-864-1116	Sequence 1116, Ap
13	949	32.1	261	10	US-09-764-864-1114	Sequence 1114, Ap
14	919	31.1	250	10	US-09-764-864-1532	Sequence 1532, Ap
15	626	21.2	171	10	US-09-764-864-1119	Sequence 1119, Ap
16	527	17.8	192	10	US-09-764-864-1117	Sequence 1117, Ap
17	474	16.0	93	10	US-09-864-761-48606	Sequence 48606, A
18	402	13.6	171	10	US-09-764-864-1536	Sequence 1536, Ap
19	238	8.1	49	10	US-09-873-637-22	Sequence 22, Appl
20	237	8.0	48	10	US-09-873-637-24	Sequence 24, Appl
21	232	7.8	49	10	US-09-873-637-18	Sequence 18, Appl
22	227	7.7	45	10	US-09-864-761-45987	Sequence 45987, A
23	227	7.7	47	10	US-09-873-637-23	Sequence 23, Appl
24	225	7.6	47	10	US-09-873-637-21	Sequence 21, Appl
25	219	7.4	48	10	US-09-873-637-20	Sequence 20, Appl
26	218	7.4	47	10	US-09-873-637-17	Sequence 17, Appl
27	190	6.4	47	10	US-09-873-637-19	Sequence 19, Appl
28	187	6.3	465	10	US-09-925-301-1198	Sequence 1198, Ap
29	171	5.8	47	10	US-09-764-864-1534	Sequence 1534, Ap
30	141	4.8	359	9	US-10-025-367-25	Sequence 25, Appl
31	139.5	4.7	1179	10	US-09-821-883-29	Sequence 29, Appl
32	135.5	4.6	359	9	US-10-025-367-23	Sequence 23, Appl
33	134.5	4.6	327	9	US-10-102-806-570	Sequence 570, App
34	133	4.5	380	9	US-10-025-367-27	Sequence 27, Appl
35	132	4.5	201	10	US-09-764-864-1113	Sequence 1113, Ap
36	131.5	4.4	279	12	US-10-014-927-19	Sequence 19, Appl
37	130	4.4	633	10	US-09-821-687-10	Sequence 10, Appl
38	129.5	4.4	545	10	US-09-978-242-1	Sequence 1, Appli
39	129.5	4.4	579	10	US-09-925-300-1415	Sequence 1415, Ap
40	128.5	4.3	561	10	US-09-821-687-4	Sequence 4, Appli
41	125.5	4.2	468	10	US-09-925-300-1620	Sequence 1620, Ap
42	124	4.2	1601	10	US-09-862-027-40	Sequence 40, Appl
43	123.5	4.2	366	9	US-10-029-180-66	Sequence 66, Appl
44	121.5	4.1	915	9	US-10-029-180-74	Sequence 74, Appl
45	120	4.1	261	9	US-10-043-487-220	Sequence 220, App

ALIGNMENTS

RESULT 1
US-09-735-705-176
; Sequence 176, Application US/09735705
; Patent No. US20020052329A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Fan, Liqun
; APPLICANT: Kalos, Michael D.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Hosken, Nancy
; APPLICANT: Fanger, Gary R.
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Henderson, Robert A.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Fanger, Neil
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY.
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C14
; CURRENT APPLICATION NUMBER: US/09/735,705
; CURRENT FILING DATE: 2000-12-12
; NUMBER OF SEQ ID NOS: 419
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 176
; LENGTH: 579
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-735-705-176

Query Match 100.0%; Score 2956; DB 10; Length 579;
Best Local Similarity 100.0%; Pred. No. 1.7e-217;
Matches 579; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGK 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGK 60

Qy      61 IELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 IELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETA 120
```

```
Qy    121 VVNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDEMAAQQNPLQQPRGRRGLGQRGSS 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 VVNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDEMAAQQNPLQQPRGRRGLGQRGSS 180

Qy    181 RQGSPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 RQGSPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAA 240

Qy    241 EKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEGRNLK 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 EKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEGRNLK 300

Qy    301 KIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIASMNL 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 KIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIASMNL 360

Qy    361 QAHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMTPPYPQFEQSETETVHQFIPALSVGAI 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 QAHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMTPPYPQFEQSETETVHQFIPALSVGAI 420

Qy    421 IGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEENFV 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 IGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEENFV 480

Qy    481 SPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVVVKIT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 SPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVVVKIT 540

Qy    541 GHFYACQVAQRKIQEILTQVKQHQQQKALQSGPPQSRRK 579
          |||||||||||||||||||||||||||||||||||||||
Db    541 GHFYACQVAQRKIQEILTQVKQHQQQKALQSGPPQSRRK 579


RESULT 2
US-09-850-716A-176
; Sequence 176, Application US/09850716A
; Patent No. US20020115139A1
; GENERAL INFORMATION:
;  APPLICANT: Kalos, Michael D.
;  APPLICANT:  McNeill, Patricia D.
;  APPLICANT:  Retter, Marc W.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE: 210121.455C15
;  CURRENT APPLICATION NUMBER: US/09/850,716A
;  CURRENT FILING DATE:  2001-05-07
;  NUMBER OF SEQ ID NOS: 440
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 176
;   LENGTH: 579
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-850-716A-176

  Query Match           100.0%;  Score 2956;  DB 10;  Length 579;
  Best Local Similarity  100.0%;  Pred. No. 1.7e-217;
  Matches  579;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGK 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGK 60

Qy     61 IELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 IELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETA 120

Qy    121 VVNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDEMAAQQNPLQQPRGRRGLGQRGSS 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 VVNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDEMAAQQNPLQQPRGRRGLGQRGSS 180

Qy    181 RQGSPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 RQGSPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAA 240

Qy    241 EKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEGRNLK 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 EKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEGRNLK 300

Qy    301 KIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIASMNL 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 KIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIASMNL 360

Qy    361 QAHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMTPPYPQFEQSETETVHQFIPALSVGAI 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 QAHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMTPPYPQFEQSETETVHQFIPALSVGAI 420

Qy    421 IGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEENFV 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 IGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEENFV 480

Qy    481 SPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVVVKIT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 SPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVVVKIT 540

Qy    541 GHFYACQVAQRKIQEILTQVKQHQQQKALQSGPPQSRRK 579
          |||||||||||||||||||||||||||||||||||||||
Db    541 GHFYACQVAQRKIQEILTQVKQHQQQKALQSGPPQSRRK 579


RESULT 3
US-09-897-778-176
; Sequence 176, Application US/09897778
; Patent No. US20020147143A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Tongtong
;  APPLICANT:  Marnerakis, Margarita
;  APPLICANT:  Fanger, Gary R.
;  APPLICANT:  Vedvick, Thomas S.
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Watanabe, Yoshihiro
;  APPLICANT:  Henderson, Robert A.
;  APPLICANT:  Peckham, David W.
;  APPLICANT:  Fanger, Neil
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE: 210121.455C16
;  CURRENT APPLICATION NUMBER: US/09/897,778
;  CURRENT FILING DATE:  2001-06-28
;  NUMBER OF SEQ ID NOS: 467
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 176
;   LENGTH: 579
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-897-778-176

  Query Match           100.0%;  Score 2956;  DB 10;  Length 579;
  Best Local Similarity  100.0%;  Pred. No. 1.7e-217;
  Matches  579;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGK 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGK 60

Qy     61 IELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 IELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETA 120

Qy    121 VVNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDEMAAQQNPLQQPRGRRGLGQRGSS 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 VVNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDEMAAQQNPLQQPRGRRGLGQRGSS 180

Qy    181 RQGSPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 RQGSPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAA 240
```

```
Qy    241 EKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEGRNLK 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 EKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEGRNLK 300

Qy    301 KIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIASMNL 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 KIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIASMNL 360

Qy    361 QAHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMTPPYPQFEQSETETVHQFIPALSVGAI 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 QAHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMTPPYPQFEQSETETVHQFIPALSVGAI 420

Qy    421 IGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEENFV 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 IGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEENFV 480

Qy    481 SPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVVVKIT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 SPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVVVKIT 540

Qy    541 GHFYACQVAQRKIQEILTQVKQHQQQKALQSGPPQSRRK 579
          |||||||||||||||||||||||||||||||||||||||
Db    541 GHFYACQVAQRKIQEILTQVKQHQQQKALQSGPPQSRRK 579


RESULT 4
US-09-735-705-348
; Sequence 348, Application US/09735705
; Patent No. US20020052329A1
; GENERAL INFORMATION:
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Fan, Liqun
;  APPLICANT:  Kalos, Michael D.
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  Hosken, Nancy
;  APPLICANT:  Fanger, Gary R.
;  APPLICANT:  Li, Samuel X.
;  APPLICANT:  Wang, Aijun
;  APPLICANT:  Skeiky, Yasir A.W.
;  APPLICANT:  Henderson, Robert A.
;  APPLICANT:  McNeill, Patricia D.
;  APPLICANT:  Fanger, Neil
;  TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE: 210121.455C14
;  CURRENT APPLICATION NUMBER: US/09/735,705
;  CURRENT FILING DATE:  2000-12-12
;  NUMBER OF SEQ ID NOS:  419
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 348
;   LENGTH: 579
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-735-705-348

  Query Match             99.6%;  Score 2943;  DB 10;  Length 579;
  Best Local Similarity   99.7%;  Pred. No. 1.7e-216;
  Matches  577;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGK 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGK 60

Qy     61 IELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 IELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETA 120

Qy    121 VVNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDEMAAQQNPLQQPRGRRGLGQRGSS 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 VVNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDETAAQQNPLQQPRGRRGLGQRGSS 180

Qy    181 RQGSPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 RQGSPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAA 240

Qy    241 EKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEGRNLK 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 EKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEGRNLK 300

Qy    301 KIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIASMNL 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 KIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIASMNL 360

Qy    361 QAHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMTPPYPQFEQSETETVHQFIPALSVGAI 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 QAHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMTPPYPQFEQSETETVHLFIPALSVGAI 420

Qy    421 IGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEENFV 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 IGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEENFV 480

Qy    481 SPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVVVKIT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 SPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVVVKIT 540

Qy    541 GHFYACQVAQRKIQEILTQVKQHQQQKALQSGPPQSRRK 579
          |||||||||||||||||||||||||||||||||||||||
Db    541 GHFYACQVAQRKIQEILTQVKQHQQQKALQSGPPQSRRK 579


RESULT 5
US-09-850-716A-348
; Sequence 348, Application US/09850716A
; Patent No. US20020115139A1
; GENERAL INFORMATION:
;  APPLICANT: Kalos, Michael D.
;  APPLICANT:  McNeill, Patricia D.
;  APPLICANT:  Retter, Marc W.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE: 210121.455C15
;  CURRENT APPLICATION NUMBER: US/09/850,716A
;  CURRENT FILING DATE:  2001-05-07
;  NUMBER OF SEQ ID NOS: 440
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 348
;   LENGTH: 579
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-850-716A-348

  Query Match             99.6%;  Score 2943;  DB 10;  Length 579;
  Best Local Similarity   99.7%;  Pred. No. 1.7e-216;
  Matches  577;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGK 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGK 60

Qy     61 IELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 IELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETA 120

Qy    121 VVNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDEMAAQQNPLQQPRGRRGLGQRGSS 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 VVNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDETAAQQNPLQQPRGRRGLGQRGSS 180

Qy    181 RQGSPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 RQGSPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAA 240

Qy    241 EKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEGRNLK 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 EKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEGRNLK 300

Qy    301 KIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIASMNL 360
```

```
Db    301 KIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIASMNL 360

Qy    361 QAHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMTPPYPQFEQSETETVHQFIPALSVGAI 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db    361 QAHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMTPPYPQFEQSETETVHLFIPALSVGAI 420

Qy    421 IGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEENFV 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 IGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEENFV 480

Qy    481 SPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVVVKIT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 SPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVVVKIT 540

Qy    541 GHFYACQVAQRKIQEILTQVKQHQQQKALQSGPPQSRRK 579
          |||||||||||||||||||||||||||||||||||||||
Db    541 GHFYACQVAQRKIQEILTQVKQHQQQKALQSGPPQSRRK 579


RESULT 6
US-09-897-778-348
; Sequence 348, Application US/09897778
; Patent No. US20020147143A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Tongtong
;  APPLICANT:  Marnerakis, Margarita
;  APPLICANT:  Fanger, Gary R.
;  APPLICANT:  Vedvick, Thomas S.
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Watanabe, Yoshihiro
;  APPLICANT:  Henderson, Robert A.
;  APPLICANT:  Peckham, David W.
;  APPLICANT:  Fanger, Neil
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE: 210121.455C16
;  CURRENT APPLICATION NUMBER: US/09/897,778
;  CURRENT FILING DATE:  2001-06-28
;  NUMBER OF SEQ ID NOS: 467
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 348
;   LENGTH: 579
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-897-778-348

  Query Match           99.6%;  Score 2943;  DB 10;  Length 579;
  Best Local Similarity  99.7%;  Pred. No. 1.7e-216;
  Matches  577;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGK 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGK 60

Qy     61 IELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 IELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETA 120

Qy    121 VVNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDEMAAQQNPLQQPRGRRGLGQRGSS 180
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db    121 VVNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDETAAQQNPLQQPRGRRGLGQRGSS 180

Qy    181 RQGSPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 RQGSPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAA 240

Qy    241 EKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEGRNLK 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 EKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEGRNLK 300

Qy    301 KIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIASMNL 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 KIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIASMNL 360

Qy    361 QAHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMTPPYPQFEQSETETVHQFIPALSVGAI 420
          |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db    361 QAHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMTPPYPQFEQSETETVHLFIPALSVGAI 420

Qy    421 IGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEENFV 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 IGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEENFV 480

Qy    481 SPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVVVKIT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 SPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVVVKIT 540

Qy    541 GHFYACQVAQRKIQEILTQVKQHQQQKALQSGPPQSRRK 579
          |||||||||||||||||||||||||||||||||||||||
Db    541 GHFYACQVAQRKIQEILTQVKQHQQQKALQSGPPQSRRK 579


RESULT 7
US-09-897-778-446
; Sequence 446, Application US/09897778
; Patent No. US20020147143A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Tongtong
;  APPLICANT:  Marnerakis, Margarita
;  APPLICANT:  Fanger, Gary R.
;  APPLICANT:  Vedvick, Thomas S.
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Watanabe, Yoshihiro
;  APPLICANT:  Henderson, Robert A.
;  APPLICANT:  Peckham, David W.
;  APPLICANT:  Fanger, Neil
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE: 210121.455C16
;  CURRENT APPLICATION NUMBER: US/09/897,778
;  CURRENT FILING DATE:  2001-06-28
;  NUMBER OF SEQ ID NOS: 467
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 446
;   LENGTH: 579
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-897-778-446

  Query Match           99.6%;  Score 2943;  DB 10;  Length 579;
  Best Local Similarity  99.7%;  Pred. No. 1.7e-216;
  Matches  577;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGK 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGK 60

Qy     61 IELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 IELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETA 120

Qy    121 VVNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDEMAAQQNPLQQPRGRRGLGQRGSS 180
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db    121 VVNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDETAAQQNPLQQPRGRRGLGQRGSS 180

Qy    181 RQGSPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 RQGSPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAA 240

Qy    241 EKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEGRNLK 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 EKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEGRNLK 300

Qy    301 KIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIASMNL 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 KIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIASMNL 360
```

```
Qy     361 QAHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMTPPYPQFEQSETETVHQFIPALSVGAI 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db     361 QAHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMTPPYPQFEQSETETVHLFIPALSVGAI 420

Qy     421 IGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEENFV 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 IGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEENFV 480

Qy     481 SPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVVVKIT 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 SPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVVVKIT 540

Qy     541 GHFYACQVAQRKIQEILTQVKQHQQQKALQSGPPQSRRK 579
           |||||||||||||||||||||||||||||||||||||||
Db     541 GHFYACQVAQRKIQEILTQVKQHQQQKALQSGPPQSRRK 579


RESULT 8
US-09-897-778-449
; Sequence 449, Application US/09897778
; Patent No. US20020147143A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Tongtong
;  APPLICANT:  Marnerakis, Margarita
;  APPLICANT:  Fanger, Gary R.
;  APPLICANT:  Vedvick, Thomas S.
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Watanabe, Yoshihiro
;  APPLICANT:  Henderson, Robert A.
;  APPLICANT:  Peckham, David W.
;  APPLICANT:  Fanger, Neil
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE: 210121.455C16
;  CURRENT APPLICATION NUMBER: US/09/897,778
;  CURRENT FILING DATE:  2001-06-28
;  NUMBER OF SEQ ID NOS: 467
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 449
;   LENGTH: 579
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-897-778-449

  Query Match             99.6%;  Score 2943;  DB 10;  Length 579;
  Best Local Similarity   99.7%;  Pred. No. 1.7e-216;
  Matches  577;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGK 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGK 60

Qy      61 IELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 IELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETA 120

Qy     121 VVNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDEMAAQQNPLQQPRGRRGLGQRGSS 180
           |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db     121 VVNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDETAAQQNPLQQPRGRRGLGQRGSS 180

Qy     181 RQGSPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAA 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 RQGSPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAA 240

Qy     241 EKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEGRNLK 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 EKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEGRNLK 300

Qy     301 KIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIASMNL 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 KIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIASMNL 360
```

```
Qy     361 QAHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMTPPYPQFEQSETETVHQFIPALSVGAI 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db     361 QAHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMTPPYPQFEQSETETVHLFIPALSVGAI 420

Qy     421 IGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEENFV 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 IGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEENFV 480

Qy     481 SPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVVVKIT 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 SPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVVVKIT 540

Qy     541 GHFYACQVAQRKIQEILTQVKQHQQQKALQSGPPQSRRK 579
           |||||||||||||||||||||||||||||||||||||||
Db     541 GHFYACQVAQRKIQEILTQVKQHQQQKALQSGPPQSRRK 579


RESULT 9
US-09-850-716A-427
; Sequence 427, Application US/09850716A
; Patent No. US20020115139A1
; GENERAL INFORMATION:
;  APPLICANT: Kalos, Michael D.
;  APPLICANT:  McNeill, Patricia D.
;  APPLICANT:  Retter, Marc W.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE: 210121.455C15
;  CURRENT APPLICATION NUMBER: US/09/850,716A
;  CURRENT FILING DATE:  2001-05-07
;  NUMBER OF SEQ ID NOS: 440
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 427
;   LENGTH: 586
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-850-716A-427

  Query Match             99.4%;  Score 2938;  DB 10;  Length 586;
  Best Local Similarity   99.7%;  Pred. No. 4.2e-216;
  Matches  576;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 NKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGKI 61
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9 NKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGKI 68

Qy      62 ELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETAV 121
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69 ELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETAV 128

Qy     122 VNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDEMAAQQNPLQQPRGRRGLGQRGSSR 181
           |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db     129 VNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDETAAQQNPLQQPRGRRGLGQRGSSR 188

Qy     182 QGSPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAAE 241
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     189 QGSPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAAE 248

Qy     242 KSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEGRNLKK 301
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     249 KSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEGRNLKK 308

Qy     302 IEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIASMNLQ 361
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     309 IEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIASMNLQ 368

Qy     362 AHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMTPPYPQFEQSETETVHQFIPALSVGAII 421
           |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db     369 AHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMTPPYPQFEQSETETVHLFIPALSVGAII 428

Qy     422 GKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEENFVS 481
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     429 GKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEENFVS 488
```

```
Qy    482 PKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVVVKITG 541
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    489 PKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVVVKITG 548

Qy    542 HFYACQVAQRKIQEILTQVKQHQQQKALQSGPPQSRRK 579
          |||||||||||||||||||||||||||||||||||||
Db    549 HFYACQVAQRKIQEILTQVKQHQQQKALQSGPPQSRRK 586


RESULT 10
US-09-897-778-427
; Sequence 427, Application US/09897778
; Patent No. US20020147143A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Tongtong
;  APPLICANT:  Marnerakis, Margarita
;  APPLICANT:  Fanger, Gary R.
;  APPLICANT:  Vedvick, Thomas S.
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Watanabe, Yoshihiro
;  APPLICANT:  Henderson, Robert A.
;  APPLICANT:  Peckham, David W.
;  APPLICANT:  Fanger, Neil
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE: 210121.455C16
;  CURRENT APPLICATION NUMBER: US/09/897,778
;  CURRENT FILING DATE:  2001-06-28
;  NUMBER OF SEQ ID NOS: 467
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 427
;   LENGTH: 586
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-897-778-427

  Query Match             99.4%;  Score 2938;  DB 10;  Length 586;
  Best Local Similarity   99.7%;  Pred. No. 4.2e-216;
  Matches  576;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 NKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGKI 61
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      9 NKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGKI 68

Qy     62 ELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETAV 121
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     69 ELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETAV 128

Qy    122 VNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDEMAAQQNPLQQPRGRRGLGQRGSSR 181
          |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    129 VNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDETAAQQNPLQQPRGRRGLGQRGSSR 188

Qy    182 QGSPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAAE 241
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    189 QGSPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAAE 248

Qy    242 KSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEGRNLKK 301
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    249 KSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEGRNLKK 308

Qy    302 IEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIASMNLQ 361
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    309 IEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIASMNLQ 368

Qy    362 AHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMTPPYPQFEQSETETVHQFIPALSVGAII 421
          ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    369 AHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMTPPYPQFEQSETETVHLFIPALSVGAII 428

Qy    422 GKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEENFVS 481
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    429 GKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEENFVS 488
```

```
Qy    482 PKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVVVKITG 541
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    489 PKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVVVKITG 548

Qy    542 HFYACQVAQRKIQEILTQVKQHQQQKALQSGPPQSRRK 579
          |||||||||||||||||||||||||||||||||||||
Db    549 HFYACQVAQRKIQEILTQVKQHQQQKALQSGPPQSRRK 586


RESULT 11
US-09-873-637-2
; Sequence 2, Application US/09873637
; Patent No. US20020061543A1
; GENERAL INFORMATION:
;  APPLICANT: Ross, Jeffrey
;  TITLE OF INVENTION: THE C-MYC CODING REGION DETERMINANT-BINDING PROTEIN
;  TITLE OF INVENTION:  (CRD-BP) AND ITS NUCLEIC ACID SEQUENCE
;  FILE REFERENCE: 960296.95131
;  CURRENT APPLICATION NUMBER: US/09/873,637
;  CURRENT FILING DATE:  2001-06-04
;  NUMBER OF SEQ ID NOS: 46
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 577
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-09-873-637-2

  Query Match             74.1%;  Score 2190;  DB 10;  Length 577;
  Best Local Similarity   74.1%;  Pred. No. 5.1e-159;
  Matches  434;  Conservative   62;  Mismatches   74;  Indels   16;  Gaps    8;

Qy      1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGK 60
          |||||||||:|:  |:||| :| : ||  || ||||:|||||||||| ||:||||  |||
Db      1 MNKLYIGNLNESVTPADLEKVFAEHKISYSGQFLVKSGYAFVDCPDEHWAMKAIETFSGK 60

Qy     61 IELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETA 120
          :|| || :|:|||||||:|| ||:|||||||  |:||||||| ||| ||:|||||:||||
Db     61 VELQGKRLEMEHSVPKKQRSRKIQIRNIPPQLRWEVLDSLLAQYGTVENCEQVNTESETA 120

Qy    121 VVNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDEMAAQQNPLQQPRGRR-GLGQRGS 179
          |||||||:::| |||: |||| |||| |||:|||||  |       ||| | | ||
Db    121 VVNVTYSNREQTRQAIMKLNGHQLENHALKVSYIPDEQITQ----GPENGRRGGFGSRGQ 176

Qy    180 SRQGSP---GSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKEN 236
          |||||   |: :||:| |:|||||||||:|||||||||||||||||||||||||||||||
Db    177 PRQGSPVAAGAPAKQQPVDIPLRLLVPTQYVGAIIGKEGATIRNITKQTQSKIDVHRKEN 236

Qy    237 AGAAEKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEG 296
          |||||:|:: ||||| |:||| |||||||||:| |  :|:|||||||||||||||||||
Db    237 AGAAEKAISVHSTPEGCSSACKMILEIMHKEAKDTKTADEVPLKILAHNNFVGRLIGKEG 296

Qy    297 RNLKKIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIA 356
          ||||:|||||:|||||| ||:|||||||||||||| :| | :||:|||||:||:||||:|
Db    297 RNLKKVEQDTETKITISSLQDLTLYNPERTITVKGAIENCCRAEQEIMKKVREAYENDVA 356

Qy    357 SMNLQAHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMT--PPYPQFEQS-ETETVHQFIP 413
          :|:||:||||||||| |:|||| :|   ||   |||::|   ||  | |: | | |||
Db    357 AMSLQSHLIPGLNLAAVGLFPASSSAVPP---PPSSVTGAAPYSSFMQAPEQEMVQVFIP 413

Qy    414 ALSVGAIIGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGK 473
          | :|||||||:|||||||||| ||||||| | ||:|||||:||||||||||||||||||
Db    414 AQAVGAIIGKKGQHIKQLSRFASASIKIAPPETPDSKVRMVVITGPPEAQFKAQGRIYGK 473

Qy    474 IKEENFVSPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDEND 533
          :|||||  |||||||| |||||: ||||||||||||||||||::|||||||||||||||||
Db    474 LKEENFFGPKEEVKLETHIRVPASAAGRVIGKGGKTVNELQNLTAAEVVVPRDQTPDEND 533

Qy    534 QVVVKITGHFYACQVAQRKIQEILTQVKQHQQQKALQSGPPQSRRK 579
          ||:||| ||||| |:|||||::|| |||| | ||  ||   |:|||
Db    534 QVIVKIIGHFYASQMAQRKIRDILAQVKQ-QHQKG-QSNLAQARRK 577
```

```
RESULT 12
US-09-764-864-1116
; Sequence 1116, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ23
;  CURRENT APPLICATION NUMBER: US/09/764,864
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 1792
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1116
;   LENGTH: 620
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (533)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1116

  Query Match             65.8%;  Score 1946.5;  DB 10;  Length 620;
  Best Local Similarity   65.1%;  Pred. No. 2.2e-140;
  Matches  385;  Conservative   81;  Mismatches   96;  Indels   29;  Gaps    8;

Qy       1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGK 60
           ||||||||||      ||  :| | |:|::|  |:|:|||||| ||::||::||| ||||
Db      23 MNKLYIGNLSPAVTADDLRQLFGDRKLPLAGQVLLKSGYAFVDYPDQNWAIRAIETLSGK 82

Qy      61 IELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETA 120
           :|||||  :||::|| |: | ||:|||||||||||||| || ||| ||: ||||||:|||
Db      83 VELHGKIMEVDYSVSKKLRSRKIQIRNIPPHLQWEVLDGLLAQYGTVENVEQVNTDTETA 142

Qy     121 VVNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDEMAAQQNPLQQPRGRRGLGQRGSS 180
           ||||||:::::|: |::||:| | ||:: |::|||||  :  :| |  | :|  |   |
Db     143 VVNVTYATREEAKIAMEKLSGHQFENYSFKISYIPDEEVSSPSPPQ--RAQR--GDHSSR 198

Qy     181 RQG-SPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGA 239
           || :||  |: :  | |||:|||||||||||||| ||:|||||||::|:|||||:||
Db     199 EQGHAPGGTSQARQIDFPLRILVPTQFVGAIIGKEGLTIKNITKQTQSRVDIHRKENSGA 258

Qy     240 AEKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEGRNL 299
           ||| :|| :|||||| ||: ||||| ||| : |  |||||||||||  ||||||||||||
Db     259 AEKPVTIHATPEGTSEACRMILEIMQKEADETKLAEEIPLKILAHNGLVGRLIGKEGRNL 318

Qy     300 KKIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIASMN 359
           |||| :| |||||| ||:|::|||||||||| || || || ||||||:||::|||: ::|
Db     319 KKIEHETGTKITISSLQDLSIYNPERTITVKGTVEACASAEIEIMKKLREAFENDMLAVN 378

Qy     360 LQAHLIPGLNLNALGLF--------PPTS--GMPPPT--------SGPPSAMTP-----P 396
           ||:|||||||:|||:|         ||    | ||          ||  |:: |     |
Db     379 QQANLIPGLNLSALGIFSTGLSVLSPPAGPRGAPPAAPYHPFTTHSGYFSSLYPHHQFGP 438

Qy     397 YPQFEQ-SETETVHQFIPALSVGAIIGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVI 455
           :|      | | |: |||  :|||||||:| |||||:||||||||||||| ||   ||||
Db     439 FPHHHSYPEQEIVNLFIPTQAVGAIIGKKGAHIKQLARFAGASIKIAPAEGPDVSERMVI 498

Qy     456 ITGPPEAQFKAQGRIYGKIKEENFVSPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQN 515
           |||||||||||||:||:||| :|||||||| ||||||  ||||||||||||||||
Db     499 ITGPPEAQFKAQGRIFGKLKEENFFNPKEEVKLEXHIRVPSSTAGRVIGKGGKTVNELQN 558

Qy     516 LSSAEVVVPRDQTPDENDQVVVKITGHFYACQVAQRKIQEILTQVKQHQQQ 566
           |:||||:||||||||||:|:|:| |||:| | |||||:||: |||| :|:
Db     559 LTSAEVIVPRDQTPDENEEVIVRIIGHFFASQTAQRKIREIVQQVKQQEQK 609


RESULT 13
US-09-764-864-1114
; Sequence 1114, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ23
;  CURRENT APPLICATION NUMBER: US/09/764,864
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 1792
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1114
;   LENGTH: 261
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-764-864-1114

  Query Match             32.1%;  Score 949;  DB 10;  Length 261;
  Best Local Similarity   74.5%;  Pred. No. 8.5e-65;
  Matches  193;  Conservative   26;  Mismatches   32;  Indels    8;  Gaps    5;

Qy     324 ERTITVKGNVETCAKAEEEIMKKIRESYENDIASMNLQAHLIPGLNLNALGLFPPTSGMP 383
           |||||||| :| | :||:|||||:||:||||:|:|:||:|||||||| |:|||| :|
Db       8 ERTITVKGAIENCCRAEQEIMKKVREAYENDVAAMSLQSHLIPGLNLAAVGLFPASSSAV 67

Qy     384 PPTSGPPSAMT--PPYPQFEQS-ETETVHQFIPALSVGAIIGKQGQHIKQLSRFAGASIK 440
           ||   |||::|   ||  | |: | | |  |||| :|||||||:|||||||||| ||||
Db      68 PP---PPSSVTGAAPYSSFMQAPEQEMVQVFIPAQAVGAIIGKKGQHIKQLSRFASASIK 124

Qy     441 IAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEENFVSPKEEVKLEAHIRVPSFAAG 500
           ||| | ||:|||||||||||||||||||||||||:|||||  |||||||| |||||: |||
Db     125 IAPPETPDSKVRMVIITGPPEAQFKAQGRIYGKLKEENFFGPKEEVKLETHIRVPASAAG 184

Qy     501 RVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVVVKITGHFYACQVAQRKIQEILTQV 560
           |||||||||||||::|||||||||||||||||:||| ||||| |:|||||::|| ||
Db     185 RVIGKGGKTVNELQNLTAAEVVVPRDQTPDENDQVIVKIIGHFYASQMAQRKIRDILAQV 244

Qy     561 KQHQQQKALQSGPPQSRRK 579
           || | ||  ||   |:|||
Db     245 KQ-QHQKG-QSNQAQARRK 261


RESULT 14
US-09-764-864-1532
; Sequence 1532, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ23
;  CURRENT APPLICATION NUMBER: US/09/764,864
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 1792
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1532
;   LENGTH: 250
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (127)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (134)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1532

  Query Match             31.1%;  Score 919;  DB 10;  Length 250;
  Best Local Similarity   73.3%;  Pred. No. 1.6e-62;
  Matches  187;  Conservative   26;  Mismatches   34;  Indels    8;  Gaps    5;

Qy     328 TVKGNVETCAKAEEEIMKKIRESYENDIASMNLQAHLIPGLNLNALGLFPPTSGMPPPTS 387
           |||| :| | :||:|||||:||:||||:|:|:||:|||||||| |:|||| :|   ||
Db       1 TVKGAIENCCRAEQEIMKKVREAYENDVAAMSLQSHLIPGLNLAAVGLFPASSSAVPP-- 58
```

```
Qy     388 GPPSAMT-PPYPQFEQS-ETETVHQFIPALSVGAIIGKQGQHIKQLSRFAGASIKIAPA 444
           |||::| || | |: | | | |||| :|||||||:|||||||||||| |||||||
Db      59 -PPSSVTGAPYSSFMQAPEQEMVQVFIPAQAVGAIIGKKGQHIKQLSRFASASIKIAPP 117

Qy     445 EAPDAKVRNVIITGPPEAQFKAQGRIYGKIKEENFVSPKEEVKLEAHIRVPSFAAGRVIG 504
           | ||:|||| |||||| |||||||||||:|||||  |||||||| |||||: |||||||
Db     118 ETPDSKVRNXIITGPPXAQFKAQGRIYGKLKEENFFGPKEEVKLETHIRVPASAAGRVIG 177

Qy     505 KGGKTVNELQNLSSAEVVVPRDQTPDENDQVVVKITGHFYACQVAQRKIQEILTQVKQHQ 564
           |||||||||||::|||||||||||||||:||| ||||| |:||||::|| |||| |
Db     178 KGGKTVNELQNLTAAEVVVPRDQTPDENDQVIVKIIGHFYASQMAQRKIRDILAQVKQ-Q 236

Qy     565 QQKALQSGPPQSRRK 579
             ||  ||  |:|||
Db     237 HQKG-QSNQAQARRK 250


RESULT 15
US-09-764-864-1119
; Sequence 1119, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ23
;  CURRENT APPLICATION NUMBER: US/09/764,864
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 1792
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1119
;   LENGTH: 171
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-764-864-1119

  Query Match             21.2%;  Score 626;  DB 10;  Length 171;
  Best Local Similarity   80.5%;  Pred. No. 2.1e-40;
  Matches  120;  Conservative   16;  Mismatches   13;  Indels    0;  Gaps    0;

Qy     418 GAIIGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEE 477
           ||||||:| |||||:|||||||||||||| ||   |||||||||||||||||||:||:|||
Db      12 GAIIGKKGAHIKQLARFAGASIKIAPAEGPDVSERMVIITGPPEAQFKAQGRIFGKLKEE 71

Qy     478 NFVSPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVVV 537
           || :|||||||||||||  |||||||||||||||||:||||:||||||||||::|:|
Db      72 NFFNPKEEVKLEAHIRVPSSTAGRVIGKGGKTVNELQNLTSAEVIVPRDQTPDENEEVIV 131

Qy     538 KITGHFYACQVAQRKIQEILTQVKQHQQQ 566
           :| |||:| | |||||:||: |||| :|:
Db     132 RIIGHFFASQTAQRKIREIVQQVKQQEQK 160


Search completed: April 24, 2003, 16:14:41
Job time : 24 secs
```

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: April 24, 2003, 16:11:08 ; Search time 21 Seconds
(without alignments)
2650.563 Million cell updates/sec

Title: US-09-897-778-176
Perfect score: 2956
Sequence: 1 MNKLYIGNLSENAAPSDLES.....VKQHQQQKALQSGPPQSRRK 579

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	418.5	14.2	854	2	T23837	hypothetical prote
2	243.5	8.2	768	2	A53184	hypothetical prote
3	240	8.1	644	2	A53184	myc far upstream e
4	236.5	8.0	568	2	T49962	hypothetical prote
5	235.5	8.0	621	2	D96554	hypothetical prote
6	231.5	7.8	313	2	T48439	probable RNA-bindi
7	229.5	7.8	398	2	T41600	probable pre-mRNA
8	227	7.7	510	2	I38489	onconeural ventral
9	225.5	7.6	589	2	T19216	hypothetical prote
10	225.5	7.6	611	2	T19217	hypothetical prote
11	222	7.5	479	2	C86275	hypothetical prote
12	219.5	7.4	396	2	S41224	hnRNP protein - Af
13	213.5	7.2	680	2	T25832	hypothetical prote
14	209.5	7.1	356	2	S58529	alpha-complex prot
15	207.5	7.0	1268	2	A44125	high density lipop
16	204.5	6.9	846	2	T04533	hypothetical prote
17	204	6.9	413	2	S46109	hnRNP complex prot
18	202.5	6.9	365	2	S42471	hnRNP protein E2 -
19	202	6.8	362	2	S78515	single-stranded nu
20	202	6.8	397	2	T30168	hypothetical prote
21	199.5	6.7	1279	2	T41389	rna binding protei
22	195	6.6	649	2	E84614	probable RNA-bindi
23	194.5	6.6	1270	2	S23464	vigilin - chicken
24	190	6.4	463	2	S41495	dC stretch-binding
25	190	6.4	464	2	S43363	transformation upr
26	190	6.4	464	2	A54143	kappa-B motif-bind
27	189.5	6.4	632	2	T02627	hypothetical prote
28	184.5	6.2	1198	2	B88279	protein C08H9.2 [i
29	184.5	6.2	1220	2	T19117	hypothetical prote
30	183	6.2	1222	2	S56030	SCP160 protein - y
31	174.5	5.9	806	2	T13690	hypothetical prote
32	172	5.8	495	2	T04255	hypothetical prote
33	170.5	5.8	629	2	T05425	polyadenylate-bind
34	166	5.6	633	1	DNHUPA	polyadenylate-bind
35	166	5.6	1283	2	T49692	related to SCP160
36	164.5	5.6	473	2	A54691	octamer-binding pr
37	161	5.4	389	2	C86460	F14M2.18 protein -
38	160.5	5.4	471	2	G01211	54 kDa protein - h
39	160.5	5.4	471	2	S41768	splicing factor ho
40	160.5	5.4	636	2	I48718	poly(A) binding pr
41	156	5.3	633	1	DNXLPA	polyadenylate-bind
42	152.5	5.2	381	2	S45766	hypothetical prote
43	152.5	5.2	638	2	S37085	polyadenylate-bind
44	152	5.1	623	2	T07933	polyadenylate-bind
45	150	5.1	344	2	S59043	spliling factor SR

ALIGNMENTS

RESULT 1
T23837
hypothetical protein M88.5 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T23837
R;Sulston, J.
submitted to the EMBL Data Library, June 1994
A;Reference number: Z19806
A;Accession: T23837
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-854 <WIL>
A;Cross-references: EMBL:Z34802; PIDN:CAA84338.1; GSPDB:GN00021; CESP:M88.5
A;Experimental source: clone M88
C;Genetics:
A;Gene: CESP:M88.5
A;Map position: 3
A;Introns: 38/3; 96/1; 133/3; 255/3; 307/3; 395/2; 436/2; 492/3; 657/3; 759/1; 820/1

Query Match 14.2%; Score 418.5; DB 2; Length 854;
Best Local Similarity 24.1%; Pred. No. 6e-19;
Matches 160; Conservative 103; Mismatches 215; Indels 187; Gaps 27;

```
Qy      78 QRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETAVVN--------VTYSSK 129
           |:::: |  |   |  : | ::  |       :|::   : | :|         |
Db     167 QQMQQQQQSNQSAHFMHQQLQAVQQQ-----QAQQMHHRLQGAPINPQQFMVPPPTMMQP 221

Qy     130 DQARQALDKLNGFQLENFTLKVAYIPDEMA--AQQ--NPLQQPRGRRGLGQRGSSRQGSP 185
           | :|| :    |:    :     |  |   ||| :| || :  :  ||   |  |
Db     222 QQMQQA-QQQQAQQMHQMQIHHQQHPQMMQQHAQQGYHPHQQNQQHQA-GQHQQSHHQSQ 279

Qy     186 G-------SVSKQKP----------C---DLPLRLLVPTQFVGAIIGKEGATIRNITKQ 224
                 : |  |           |   | |:| :|  ::   |||  |:||::|
Db     280 NHNQHRNHNQSHSGPHHIPQNLMMPRCMLKDWPIRCVVEGKYHAVIIGPNGSTIKDIASS 339

Qy     225 TQSKID---VHRKENA--GAAEKSITILSTPEGTSAACKSILEIMHKEA--QDIKFTEEI 277
           |: ::|   : :||    |  :: :|:    |  : |   ||::: ||   |:    :
Db     340 TRCRVDFVNLSKKERTVLGNNDRILTVHGVAEQATKAVARILDVIQSEAVKDDVNVGADT 399

Qy     278 PLKILAHNNFVGRLIGKEGRNLKKIEQDTDTKITISPLQE-------LT----LYNPERT 326
           |:: |||   ||||| | ::|:| | | | ||::   |       ||    |   |||
Db     400 VLRMRAHNQLCGRLIGKAGSSIKEIMQKTGTNITVTKYIEPPGGISGLTANELLGLMERT 459

Qy     327 ITVKG-NVETCAKAEEEIMKKIRESYENDIASMNLQAHLIPGLNLNALGLFPPTSGMPP- 384
           | |:| ::|   :||  |  |::: ||:| : :  |:   |         ||   |||
Db     460 IMVRGPSIEAVVQAEAALISAKLKKCYESD-SQLRAQSMQCP---------MPPMM-MPPI 508

Qy     385 -PTSGPPSAMTPPY----------PQFEQSETETVHQ---------------------- 410
           |     ||:: |:           |   |    |||
Db     509 LPPGASSSAVSAPHFIPTPVGVMQVQHFASSQHLVHQNANNSFLQPGVLQIQPGTTNLRQ 568
```

```
Qy    411 ---FIPALSVGAIIGKQGQHIKQLSRFAGASIKIAPAEAPDAKV---------------- 451
Db    569 VRMWVPDSMIGALIGAKGKNIKMIIRDTGASVKI---EAPEEKTQREAEEAEKKRKLDET 625

Qy    452 ------------------------------------RMVIITGPPEAQFKAQGRIYGK 473
Db    626 DSGCEGVASGDHPQEFLEDNATINSSDAIEEKPKPVSERMVTINGDDLQLLKAQSYVFSK 685

Qy    474 IKEENFVSP--------KEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPR 525
Db    686 IAETSSSLPSSGMDGDRSHMLRIRTEVSVPTRIIGRIIGKGGQNVRELQRITGAVVKIPE 745

Qy    526 DQTPD----------ENDQVVVKITGHFYACQVAQRKIQEILTQVKQ---HQQQKA-LQS 571
Db    746 EERNGGEVYRHDDGLEEDMTMIRTIGNMYSTHNVQFRLAHLVNEYYRSGDHRNKSSDYKG 805

Qy    572 GPPQS 576
Db    806 GRPHS 810
```

RESULT 2
T27855
hypothetical protein ZK418.9 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T27855
R;Fulton, L.
submitted to the EMBL Data Library, April 1994
A;Description: The sequence of C. elegans cosmid ZK418.
A;Reference number: Z20430
A;Accession: T27855
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-768 <FUL>
A;Cross-references: EMBL:U00047; PIDN:AAA50693.1; CESP:ZK418.9
A;Experimental source: strain Bristol N2
C;Genetics:
A;Gene: CESP:ZK418.9
A;Introns: 166/1; 221/3; 243/1; 279/1; 387/3; 490/1; 567/3; 727/2

  Query Match 8.2%; Score 243.5; DB 2; Length 768;
  Best Local Similarity 21.2%; Pred. No. 9.2e-08;
  Matches 123; Conservative 103; Mismatches 215; Indels 139; Gaps 22;

```
Qy     62 ELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETAV 121
Db    139 QLIGKLIELD--ITKNESL--------------LLSALFL-------CSAVHPDISTEG 174

Qy    122 VNVTYSSKDQARQAL----------------DKLNGFQ-LENFTLKVAYIPDEMAAQQN 163
Db    175 KNLLYKYQQYYSAALINHCCLTNQQNAPSRYSELLSLYQIIEETHQKIIGADGGIGTLKR 234

Qy    164 PLQQ--------PRGRRGLGQRGSSRQGSPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEG 215
Db    235 PLDSEILDGDLIPTKKS--SEVGDLNMGDSDKITDIYP--------VPEKVVGLVIGKGG 284

Qy    216 ATIRNITKQTQSKIDVHRKENAGAAEKSITILSTPEGTSAACKSILEIMHKE---AQDIK 272
Db    285 SEIRLIQQTSGCRVQMDPDHQSVNGFRNCTIEGPPDQVAVARQMITQVINRNQTGAQPGA 344

Qy    273 FTEEIPLKILAHNNFVGRLIGKEGRNLKKIEQDTDTKITISPLQELTLY----NPERTIT 328
Db    345 APGEVTEEMLIPADKIGLVIGKGGETIRIVQEQSGLR-NCNVVQETTTATGQPKPLRMIG 403

Qy    329 VKGNVETCAKAEEEIMKKIRESYENDIASMNLQAHLIPGLNLNALGLFPPTSGMPPPTSG 388
Db    404 SPAAIET-AKA----------------------LVHNIMNNTQGNAPLLQRAPHQPSG 438

Qy    389 PPSAMTPPYPQF--------EQSETETVHQFIPALSVGAIIGKQGQHIKQLSRFAGASIK 440
Db    439 ----------QFGGGYGAQEAQAKGEVI---VPRLSAGMIIGKGGEMIKRLAAETGTKIQ 485

Qy    441 IAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKE--ENFVSPKEEVKLEA-------- 490
Db    486 FKPDTNPNSEDRIAVIMGTRDQIYRATERITEIVNRAIKNNGAPQDRGSAGTVLPGQSIF 545

Qy    491 HIRVPSFAAGRVIGKGGKTVNELQNLSSAEV-VVPRDQTPDENDQVVVKITGH----FYA 545
Db    546 YMHVPAGKCGLVIGKGGENIKQIERETGATCGLAPAAEQKNE-DEKVFEIKGSQLQIHHA 604

Qy    546 CQVAQRKIQEILTQVK---------QHQQQKALQSGPPQS 576
Db    605 SHLVRIKVGEISPNTPVPPLQGAGGGYQQQQAMFSAGTQN 644
```

RESULT 3
A53184
myc far upstream element-binding protein - human
N;Alternate names: FUSE-binding protein
C;Species: Homo sapiens (man)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 05-Nov-1999
C;Accession: A53184
R;Duncan, R.; Bazar, L.; Michelotti, G.; Tomonaga, T.; Krutzsch, H.; Avigan, M.; Leve
Genes Dev. 8, 465-480, 1994
A;Title: A sequence-specific, single-strand binding protein activates the far upstrea
A;Reference number: A53184; MUID:94170991; PMID:8125259
A;Accession: A53184
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-644 <DUN>
A;Cross-references: GB:U05040; NID:g460151; PIDN:AAA17976.1; PID:g460152
C;Keywords: DNA binding

  Query Match 8.1%; Score 240; DB 2; Length 644;
  Best Local Similarity 22.3%; Pred. No. 1.2e-07;
  Matches 100; Conservative 75; Mismatches 154; Indels 120; Gaps 17;

```
Qy    161 QQNPLQ---QPRGRRGLGQRGSSRQGSPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGAT 217
Db     63 QKRPLEDGDQPDAKKVAPQNDSFGTQLPPMHQQQSRSVMTEEYKVPDGMVGFIIGRGGEQ 122

Qy    218 IRNITKQTQSKIDVHRKENAGAAEKSITILSTPEGTSAACKSILEIMHKEAQDIKFTE-- 275
Db    123 ISRIQQESGCKIQI-APDSGGLPERSCMLTGTPESVQSAKRLLDQIVEKGRPAPGFHHGD 181

Qy    276 ---EIPLKILAHNNFVGRLIGKEGRNLKKIEQDTDTKITISPLQELTLYNPERT-----I 327
Db    182 GPGNAVQEIMIPASKAGLVIGKGGETIKQLQERAGVKMVM--IQD----GPQNTGADKPL 235

Qy    328 TVKGNVETCAKAEEEIMKKIRE-----SYENDIASMNLQAHLIPGLNLNALGLFPPTSGM 382
Db    236 RITGDPYKVQQAKEMVLELIRDQGGFREVRNEYGSR------IGG----------NEGI 278

Qy    383 PPPTSGPPSAMTPPYPQFEQSETETVHQFIPALSVGAIIGKQGQHIKQLSRFAGASIKIA 442
Db    279 DVP-------------------------IPRFAVGIVIGRNGEMIKKIQNDAGVRIQFK 312

Qy    443 PAEAPDAKVRMVIITGPPE-AQFKAQ-------------------GRIYGK-------- 473
Db    313 PDDGTTPE-RIAQITGPPDRCQHAAEIITDLLRSVQAGNPGGPGPGGRGRGRGQGNWNMG 371

Qy    474 ----IKEENFVSPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTP 529
Db    372 PPGGLQEFNFI-------------VPTGKTGLIIGKGGETIKSISQQSGARIELQRNPPP 418

Qy    530 DENDQV-VVKITGH----FYACQVAQRKI 553
Db    419 NADPNMKLFTIRGTPQQIDYARQLIEEKI 447
```

RESULT 4
T49962
hypothetical protein F8M21.160 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000

```
C;Accession: T49962
R;Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Rudd, S.; Le
submitted to the Protein Sequence Database, April 2000
A;Reference number: Z24493
A;Accession: T49962
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-568 <BEV>
A;Cross-references: EMBL:AL353993; GSPDB:GN00063; ATSP:F8M21.160
A;Experimental source: cultivar Columbia; BAC clone F8M21
C;Genetics:
A;Gene: ATSP:F8M21.160
A;Map position: 5
A;Introns: 200/3; 337/3; 544/3

  Query Match             8.0%;  Score 236.5;  DB 2;  Length 568;
  Best Local Similarity   22.8%;  Pred. No. 1.7e-07;
  Matches   92;  Conservative   79;  Mismatches  144;  Indels   89;  Gaps   16;

Qy    195 DLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAAEKSITILSTPEGTS 254
          |   | | | : :|::||: |  :: :   |:||| :  :   |  |: ||| |  : |:
Db     49 DTVFRYLCPVKKIGSVIGRGGDIVKQLRNDTRSKIRI-GEAIPGCDERVITIYSPSDETN 107

Qy    255 A----------ACKSILEIMHKEAQDIKFTEEIP-------LKILAHNNFVGRLIGKEGR 297
          |          |  ::  |  :  |   :|: |        |:|  :: :| ::|: |:
Db    108 AFGDGEKVLSPAQDALFRIHDRVVADDARSEDSPEGEKQVTAKLLVPSDQIGCILGRGGQ 167

Qy    298 NLKKIEQDTDTKITISPLQELTL--YNPERTITVKGNVETCAKAEEEIMKKIRESYENDI 355
          :: |  :|  :| |   : : |   | :  | : | |    || :|  ::   :||
Db    168 IVQNIRSETGAQIRIVKDRNMPLCALNSDELIQISGEVLIVKKALLQIASRL---HENPS 224

Qy    356 ASMNLQAHLIPGLNLNALGLFPPTS----------GMPP-------------PTSGPP 390
          | | ||         |:: | :|  |           |: |             |   ||
Db    225 RSQNL---------LSSSGGYPAGSLMSHAGGPRLVGLAPLMGSYGRDAGDWSRPLYQPP 275

Qy    391 SAMTPPYPQFEQSETETVHQFIPALSVGAIIGKQGQHIKQLSRFAGASIKI--APAEAPD 448
             ||  :|       :    |  :: ::||| |  | || :   |:||:  :  |  |
Db    276 -RNDPPATEF------FIRLVSPVENIASVIGKGGALINQLRQETRATIKVDSSRTEGND 328

Qy    449 AKVRMVIITGPP----------EAQFKAQGRIYGKIKEENFVSPKEEVKLEAHIRVPSFA 498
             :: |:           ||  : | :   |:: :: :     |     : |||
Db    329 C---LITISAREVFEDAYSPTIEAVMRLQPKCSDKVERDSGL-----VSFTTRLLVPSSR 380

Qy    499 AGRVIGKGGKTVNELQNLSSAEV-VVPRDQTP----DENDQVVV 537
          | ::||||  : |:: :: | : :: ::   |    |::: | |
Db    381 IGCILGKGGAIITEMRRMTKANIRILGKENLPKVASDDDEMVQV 424


RESULT 5
D96554
hypothetical protein F19C24.19 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: D96554
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: D96554
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-621 <STO>
A;Cross-references: GB:AE005173; NID:g11094762; PIDN:AAG29695.1; GSPDB:GN00141
C;Genetics:
A;Gene: F19C24.19
A;Map position: 1

  Query Match             8.0%;  Score 235.5;  DB 2;  Length 621;
  Best Local Similarity   22.9%;  Pred. No. 2.2e-07;
  Matches  104;  Conservative   76;  Mismatches  171;  Indels  103;  Gaps   17;

Qy    199 RLLVPTQFVGAIIGKEGATIRNITKQTQSKI----------------------------D 230
          ||| |    |||||| |: ||::   | |||                           :
Db     22 RLLCPATRTGAIIGKGGSVIRHLQSVTGSKIRVIDDIPVPSEERVVLIIAPSGKKKDESN 81

Qy    231 VHRKENAGA----AEKSITILST-----PEGTSAACKSILEIMHK-----EAQDI----- 271
          |   || |:     ||       |       |  |:|  ::| :  :      :|  :
Db     82 VCDSENPGSEEPKQEKGSECAGTSGGDDEEAPSSAQMALLRVFERIVFGDDAATVDGDEL 141

Qy    272 -KFTEEIPLKILAHNNFVGRLIGKEGRNLKKIEQDTDTKITISPLQEL--TLYNPERTIT 328
           |   |   :::   | |  |: | |: ::|| :|:   : ||   ::    :  :  |
Db    142 DKGESEGLCRMIVRGNQVDYLMSKGGKMIQKIREDSGAIVRISSTDQIPPCAFPGDVVIQ 201

Qy    329 VKGNVETCAKAEEEIMKKIRES--------------YENDIASMNLQAHLIPGLNLNAL 373
          : |   :  ||   :   ::||              |  :  ||
Db    202 MNGKFSSVKKALLLVTNCLQESGAPPTWDECPFPQPGYPPEYHSME-------------- 247

Qy    374 GLFPPTSGMPPPTSGPPSAMTPPYPQFEQSETETVHQFIPALSVGAIIGKQGQHIKQLSR 433
            : |    |||   |      |: :    |        ||  ||::||| |  :: |
Db    248 --YHPQWDHPPPNPMPED--VGPFNRPVVEEEVAFRLLCPADKVGSLIGKGGAVVRALQN 303

Qy    434 FAGASIKIAPAEAPDAKVRMVIITGPPEAQFK---AQG---RIYGKIKEENFVSPKEEVK 487
          :||||::     |:: |:::|:     : :   ||    |:: :| |  |   :
Db    304 ESGASIKVSD-PTHDSEERIIVISARENLERRHSLAQDGVMRVHNRIVEIGF---EPSAA 359

Qy    488 LEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEV-VVPRDQTPDENDQ--VVVKITGHFY 544
          : | : | |   ||::|||| ::|::  : | : |  :||       |   :|:: |:
Db    360 VVARLLVHSPYIGRLLGKGGHLISEMRRATGASIRVFAKDQATKYESQHDEIVQVIGNL- 418

Qy    545 ACQVAQRKIQEILTQVKQHQQQKAL----QSGPP 574
          :  |  : :|| ::::      |     |||
Db    419 --KTVQDALFQILCRLREAMFPGRLPFQGMGGPP 450


RESULT 6
T48439
probable RNA-binding protein - Arabidopsis thaliana
N;Alternate names: protein T32M21.30
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C;Accession: T48439
R;Bevan, M.; Terryn, N.; Ardiles, W.; Buysshaert, C.; Dasseville, R.; De Clerck, R ,
ewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.
submitted to the Protein Sequence Database, March 2000
A;Reference number: Z24487
A;Accession: T48439
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-313 <BEV>
A;Cross-references: EMBL:AL162875
A;Experimental source: cultivar Columbia; BAC clone T32M21
C;Genetics:
A;Map position: 5
A;Introns: 29/1; 108/3; 144/2; 209/1; 229/3; 261/3; 282/2
A;Note: T32M21.30

  Query Match             7.8%;  Score 231.5;  DB 2;  Length 313;
  Best Local Similarity   25.1%;  Pred. No. 1.4e-07;
  Matches   79;  Conservative   62;  Mismatches  131;  Indels   43;  Gaps   10;

Qy    183 GSPGSVSKQKP----------CDLP--LRLLVPTQFVGAIIGKEGATIRNITKQTQSKID 230
          |||  ::|: |           : |  :| ||     |::||| |:|     :: ::|
Db     10 GSPEELAKRSPEPHDSSEADSAEKPTHIRFLVSNAAAGSVIGKGGSTITEFQAKSGARIQ 69

Qy    231 VHRKEN--AGAAEKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIP---LKILAHN 285
          : | :    |  :: | |  : :      : ||: :| |       |  |   :::: |
Db     70 LSRNQEFFPGTTDRIIMISGSIKEVVNGLELILDKLHSELHAEDGNEVEPRRRIRLVVPN 129

Qy    286 NFVGRLIGKEGRNLKKIEQDTDTKITISPLQELTLYN-PERTITVKGNVETCAKAEEEIM 344
```

```
Db    130 SSCGGIIGKGGATIKSFIEESKAGIKISPLDN-TFYGLSDRLVTLSGTFEEQMRAIDLIL 188

Qy    345 KKIRESYENDIASMNLQAHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMTPPYPQFEQSE 404

Db    189 AKLT---EDDHYSQNVHSPYSYAAGYNSVNYAPNGSG----------GKYQNHKEEA 232

Qy    405 TETVHQFIPALSVGAIIGKQGQHIKQLSRFAGASIKIAP-----AEAPDAKVRMVIITGP 459

Db    233 STTVTIGVADEHIGLVLGRGGRNIMEITQMTGARIKISDRGDFMSGTTDRKVS---ITGP 289

Qy    460 PEAQFKAQGRIYGKI 474

Db    290 QRAIQQAETMIKQKV 304


RESULT 7
T41600
probable pre-mRNA splicing factor - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 01-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T41600
R;Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Bothe, G.; Pohl, T.
submitted to the EMBL Data Library, October 1998
A;Reference number: Z22003
A;Accession: T41600
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-398 <LYN>
A;Cross-references: EMBL:AL031825; PIDN:CAA21234.1; GSPDB:GN00068; SPDB:SPCC757.09c
A;Experimental source: strain 972h-; cosmid c757
C;Genetics:
A;Gene: SPDB:SPCC757.09c
A;Map position: 3

  Query Match             7.8%;  Score 229.5;  DB 2;  Length 398;
  Best Local Similarity   26.5%;  Pred. No. 2.8e-07;
  Matches   90;  Conservative   55;  Mismatches  148;  Indels   47;  Gaps   11;

Qy    176 QRGSSRQGSPGS----VSKQKPCD--------LPLRLLVPTQFVGAIIGKEGATIRNITK 223

Db     62 QQSMTNQPEPTSQVPPISAKPPMDDATYATQQLTLRALLSTREAGIIIGKAGKNVAELRS 121

Qy    224 QTQSKIDVHRKENAGAAEKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEE-----IP 278

Db    122 TTNVKAGV-TKAVPNVHDRVLTISGPLENVVRAYRFIIDIFAKNSTNPDGTPSDANTPRK 180

Qy    279 LKILAHNNFVGRLIGKEGRNLKKIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAK 338

Db    181 LRLLIAHSLMGSIIGRNGLRIKLIQDKCSCRMIAS--KDMLPQSTERTVEIHGTVDNLHA 238

Qy    339 AEEEIMKKIRESYENDIASM--NLQAHL---IPGLNLNA--LGLFPPTSGMPPPTSGP-- 389

Db    239 AIWEIGKCLIDDWERGAGTVFYNPVSRLTQPLPSLASTASPQQVSPPAA--PSTTSGEAI 296

Qy    390 ----------PSAMTPPYPQFEQSETETVHQFIPALSVGAIIGKQGQHIKQLSRFAGA 437

Db    297 PENFVSYGAQVFPATQMP----FLQQPKVTQNISIPADMVGCIIGRGGSKISEIRRTSGS 352

Qy    438 SIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEE 477

Db    353 KISIAKEPHDETGERMFTITGTHEENEKALFLLYQQLEME 392


RESULT 8
I38489
onconeural ventral antigen-1 - human
C;Species: Homo sapiens (man)
C;Date: 06-Sep-1996 #sequence_revision 06-Sep-1996 #text_change 05-Nov-1999
C;Accession: I38489
R;Buckanovich, R.J.; Posner, J.B.; Darnell, R.B.
Neuron 11, 657-672, 1993
A;Title: Nova, the Paraneoplastic Ri Antigen, is Homologous to an RNA-Binding Protein an
A;Reference number: I38489; MUID:94000830; PMID:8398153
A;Accession: I38489
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-510 <RES>
A;Cross-references: EMBL:U04840; NID:g440877; PIDN:AAA16022.1; PID:g440878
C;Genetics:
A;Gene: Nova-1

  Query Match             7.7%;  Score 227;  DB 2;  Length 510;
  Best Local Similarity   22.4%;  Pred. No. 5.8e-07;
  Matches  119;  Conservative   77;  Mismatches  179;  Indels  156;  Gaps   20;

Qy    155 PDEMAAQQNPLQQPRGRRGLGQRGSSRQGSPGSVSKQKPCDLPLRLLVPTQFVGAIIGKE 214

Db     23 PD---SRKRPLEAP------PEAGSTKRTNTGEDGQYF-----LKVLIPSYAAGSIIGKG 68

Qy    215 GATIRNITKQTQSKIDVHRKENA-----GAAEKSITILSTPEGTSAACKSILEIMHKEAQ 269

Db     69 GQTIVQLQKETGATIKLSKLSKSKDFYPGTTERVCLIQGTVEALNAVHGFIAEKIREMPQ 128

Qy    270 DIKFTEEIPL----------------------------------------KILAHNNFVGR 290

Db    129 NVAKTEPVSILQPQTTVNPDRIKQTLPSSPTTTKSSPSDPMTTSRANQVKIIVPNSTAGL 188

Qy    291 LIGKEGRNLKKIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRE- 349

Db    189 IIGKGGATVKAVMEQSGAWVQLSQKPD-GINLQERVVTVSGEPEQNRKAVELIIQKIQED 247

Qy    350 ---------SYEN---DIASMN----------------LQAHLIPGLNLNALGLFPPT- 379

Db    248 PQSGSCLNISYANVTGPVANSNPTGSPYANTAEVLPTAAAAAGLLGHANLAGVAAFPAVL 307

Qy    380 SGMPPPTSGPPSAMTPPYPQFEQSETETVHQFIPALSVGAIIGKQGQHIKQLSRFAGASI 439

Db    308 SGF---TGNDLVAIT--------SALNTL------ASYGYNLNTLGLGLSQAAA-TGALA 349

Qy    440 KIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKI------------------------- 474

Db    350 AAAASANPAAAAANLLATYASEA--SASGSTAGGTAGTFALGSLAAATAATNGYFGAASP 407

Qy    475 --------KEENFVSPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPR- 525

Db    408 LAASAILGTEKSTDGSKDVVE----IAVPENLVGAILGKGGKTLVEYQELTGARIQISKK 463

Qy    526 -DQTPDENDQVVVKITGHFYACQVAQRKIQEILTQVKQHQQQKALQSGPPQ 575

Db    464 GEFVPGTRNRKVT-ITGTPAATQAAQYLITQRIT------YEQGVRAANPQ 507


RESULT 9
T19216
hypothetical protein C12D8.1a - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T19216
R;McMurray, A.
submitted to the EMBL Data Library, June 1996
A;Reference number: Z19092
A;Accession: T19216
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-589 <WIL>
A;Cross-references: EMBL:Z73969; PIDN:CAA98232.1; GSPDB:GN00023; CESP:C12D8.1a
A;Experimental source: clone C12D8
C;Genetics:
A;Gene: CESP:C12D8.1a
A;Map position: 5
A;Introns: 7/3; 25/1; 159/3; 318/3; 513/2

  Query Match             7.6%;  Score 225.5;  DB 2;  Length 589;
  Best Local Similarity   23.4%;  Pred. No. 8.9e-07;
  Matches   86;  Conservative   72;  Mismatches  157;  Indels   53;  Gaps   11;
```

```
Qy    202 VPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAAEKSITILSTPEGTSAACKSIL 261
Db     54 IPESAVGIVIGRGGSEIQGIQAKAGCRVQMSPDADPSSGVRMVTLEGSRSNVETAKHLIN 113

Qy    262 EIMHKEAQDIKFTEEIP-----LKILAHNNFVGRLIGKEGRNLKKIEQDTDTK-ITISPL 315
Db    114 EVVAR-SQNPRPQYGFPRAQTTIDIAIPPNRCGLIIGKSGDTIRQLQEKSGCKMILVQDN 172

Qy    316 QELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIASMNLQAHLIPGLNLNALGL 375
Db    173 QSVS--DQSKPLRITGDPQKIELA-KQLVAEILNSGGDGNGGSGLQMHHAGG-------- 221

Qy    376 FPPTSGMPPPTSGPPSAMTPPYPQFEQSETETVHQFIPALSVGAIIGKQGQHIKQLSRFA 435
Db    222 ----------GGGASA----------RGEVV---VPRSSVGIIIGKQGDTIKRLAMET 256

Qy    436 GASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEENFVS------PKEEVKLE 489
Db    257 GTKIQFKPDDDPSTPERCAVIMGTRDQIYRATERITELVKKSTMQQGGGGNVAGAMVSNE 316

Qy    490 A---HIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVVVKITGHFYAC 546
Db    317 ASTFYMSVPAAKCGLVIGKGGETIKQINSESGAHCELSRDPTGNADEKVFV-IKGGKRAI 375

Qy    547 QVAQRKIQ 554
Db    376 EHAKHLIR 383
```

RESULT 10
T19217
hypothetical protein C12D8.1b - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T19217
R;McMurray, A.
submitted to the EMBL Data Library, June 1996
A;Reference number: Z19092
A;Accession: T19217
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-611 <WIL>
A;Cross-references: EMBL:Z73969; PIDN:CAA98233.1; GSPDB:GN00023; CESP:C12D8.1b
A;Experimental source: clone C12D8
C;Genetics:
A;Gene: CESP:C12D8.1b
A;Map position: 5
A;Introns: 181/3; 340/3; 535/2

  Query Match 7.6%; Score 225.5; DB 2; Length 611;
  Best Local Similarity 23.4%; Pred. No. 9.4e-07;
  Matches 86; Conservative 72; Mismatches 157; Indels 53; Gaps 11;

```
Qy    202 VPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAAEKSITILSTPEGTSAACKSIL 261
Db     76 IPESAVGIVIGRGGSEIQGIQAKAGCRVQMSPDADPSSGVRMVTLEGSRSNVETAKHLIN 135

Qy    262 EIMHKEAQDIKFTEEIP-----LKILAHNNFVGRLIGKEGRNLKKIEQDTDTK-ITISPL 315
Db    136 EVVAR-SQNPRPQYGFPRAQTTIDIAIPPNRCGLIIGKSGDTIRQLQEKSGCKMILVQDN 194

Qy    316 QELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIASMNLQAHLIPGLNLNALGL 375
Db    195 QSVS--DQSKPLRITGDPQKIELA-KQLVAEILNSGGDGNGGSGLQMHHAGG-------- 243

Qy    376 FPPTSGMPPPTSGPPSAMTPPYPQFEQSETETVHQFIPALSVGAIIGKQGQHIKQLSRFA 435
Db    244 ----------GGGASA----------RGEVV---VPRSSVGIIIGKQGDTIKRLAMET 278

Qy    436 GASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEENFVS------PKEEVKLE 489
Db    279 GTKIQFKPDDDPSTPERCAVIMGTRDQIYRATERITELVKKSTMQQGGGGNVAGAMVSNE 338

Qy    490 A---HIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVVVKITGHFYAC 546
Db    339 ASTFYMSVPAAKCGLVIGKGGETIKQINSESGAHCELSRDPTGNADEKVFV-IKGGKRAI 397

Qy    547 QVAQRKIQ 554
Db    398 EHAKHLIR 405
```

RESULT 11
C86275
hypothetical protein [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001
C;Accession: C86275
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: C86275
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-479 <STO>
A;Cross-references: GB:AE005172; NID:g5080792; PIDN:AAD39302.1; GSPDB:GN00141
C;Genetics:
A;Map position: 1

  Query Match 7.5%; Score 222; DB 2; Length 479;
  Best Local Similarity 21.6%; Pred. No. 1.1e-06;
  Matches 98; Conservative 93; Mismatches 169; Indels 94; Gaps 19;

```
Qy    165 LQQPRGRRG------LGQRGSSRQGSPGSVSKQKPC--DLPLRLLVPTQFVGAIIGKEGA 216
Db      5 LRNIHGKRSNLQSEFTGNGGSKRRNLHDETDQNVIASEDTVYRYLCPVKKTGSIIGKGGE 64

Qy    217 TIRNITKQTQSKIDVHRKENAGAAEKSITILSTPEGTS-------AACKSILEIMHK--- 266
Db     65 IAKQIRSETKSNMRIN-EALPGCEERVVTMYSTNEELNHFGDDGELVCPA-LDALFKVHD 122

Qy    267 -------------EAQDIKFTEEIPLKILAHNNFVGRLIGKEGRNLKKIEQDTDTKITI- 312
Db    123 MVVADADQDDGTDDDNDLGEKQTVTVRMLVPSDQIGCVIGKGGQVIQNLRNDTNAQIRVI 182

Qy    313 ---SPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRES---YENDIASMNLQAHLIP 366
Db    183 KDHLPACALTLSHDELLLII-GEPLVVREALYQVASLLHDNPSRFQHLLLSSSSSSMHQP 241

Qy    367 GLNLNALGLFPPTSGMPPPTSGPPSAMTPPYPQFE-QSETETVHQF-----IPALSVGAI 420
Db    242 GAMLMS-----------------AALTSSHRNYAVRRDIADAREFCVCFICPAENVGGV 283

Qy    421 IGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPE----------AQFKAQGRI 470
Db    284 IGKGGGFINQIRQETGATIRVNTSETDDDD--CIIFISSKEFYEDQSPAVNAAIRLQQRC 341

Qy    471 YGKI-KEENFVSPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEV-VVPRDQT 528
Db    342 SEKVGKDAN------DLAISTRLLVSSSQIGCLIGKGGAVISEMRSVTRANIRILQKEDV 395

Qy    529 PD--ENDQVVVKITGHFYACQVAQRKIQEILTQV 560
Db    396 PKIAREDEEMVQITGSPDAAMKA-------LTQV 422
```

RESULT 12

```
S41224
hnRNP protein - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 19-May-1994 #sequence_revision 10-Nov-1995 #text_change 20-Jun-2000
C;Accession: S41224
R;Siomi, H.; Matunis, M.J.; Michael, W.M.; Dreyfuss, G.
Nucleic Acids Res. 21, 1193-1198, 1993
A;Title: The pre-mRNA binding K protein contains a novel evolutionarily conserved motif.
A;Reference number: S41224; MUID:93219080; PMID:8464704
A;Accession: S41224
A;Molecule type: mRNA
A;Residues: 1-396 <SIO>
C;Superfamily: transformation upregulated nuclear protein HNRPK

  Query Match             7.4%;  Score 219.5;  DB 2;  Length 396;
  Best Local Similarity   21.3%;  Pred. No. 1.2e-06;
  Matches   93;  Conservative   76;  Mismatches  166;  Indels  101;  Gaps   18;

Qy     175 GQRGSSRQGSPGSVSKQKPCDL--PLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVH 232
           |:| :       :  : :  |:   ||:|: ::  ||:||| |  |: :     : : |
Db      17 GKRPAEDMEDEQAFKRSRNTDMMVELRILLQSKNAGAVIGKGGKNIKALRTDYNASVSV- 75

Qy     233 RKENAGAAEKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLI 292
             ::   |: ::| :  |      | |:  : :  :   |  :  |::| | :  | :|
Db      76 --PDSSGPERILSISADIETIGEILKKIIPTLEEHFKGNDF--DCELRLLIHQSLAGGII 131

Qy     293 GKEGRNLKKIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRES-- 350
           | :|  :|:: : | |  ||   ||   :: :| : : |  :   :  : |:  | ||
Db     132 GVKGAKIKELREKTQT--TIKLFQECCPHSTDRVVLIGGRPDRVVECIKVILDLISESPV 189

Qy     351 -----------YEN-DIASMNL-------QAHLIPGLNLNALGLF----PPTSGMPPPTS 387
                      ||  |     :       : |   | :::| | |    |   | | | |
Db     190 KGRSQPYDPNFYETYDYGGFTMMFDDRRGRPH---GFSMHARGGFDRMPPGPGGRPMPQS 246

Qy     388 -------GPPSAMTPPYPQFEQSETETVHQFIPALSVGAIIGKQGQHIKQLSRFAGAS-- 438
                 |     || |                        |: |:          | |
Db     247 RRDYDDMSPRRGPLPPPP----------------------GRGGR---------GGSRA 274

Qy     439 --IKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEENFVSPKEEVKLEAHIRVPS 496
             : : |   |    |     | |:      || ||:    :   |     :   : :|
Db     275 RNLPLPPPPPPRGGDR----RGRPDHYDGMGGRGYGRGSFGDIGGP----VITTQVTIPK 326

Qy     497 FAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVVVKITGHFYACQVAQRKIQE- 555
             || :|||||: : :::: | | : :  |:  : :|  :: |||        | :||
Db     327 DLAGSIIGKGGQRIKQIRHESGASIKI--DEPLEGSDDRIITITG-------TQDQIQNA 377

Qy     556 ---ILTQVKQHQQQKA 568
               :   |||  :  |
Db     378 RFLLQNSVKQFSEDYA 393


RESULT 13
T25832
hypothetical protein M01A10.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T25832
R;Scheet, P.
submitted to the EMBL Data Library, February 1997
A;Description: The sequence of C. elegans cosmid M01A10.
A;Reference number: Z20094
A;Accession: T25832
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-680 <SCH>
A;Cross-references: EMBL:U88174; PIDN:AAB42272.1; GSPDB:GN00019; CESP:M01A10.1
A;Experimental source: strain Bristol N2; clone M01A10
C;Genetics:
A;Gene: CESP:M01A10.1
A;Map position: 1
A;Introns: 63/2; 97/3; 121/1; 160/1; 269/3; 411/3; 512/1; 649/2

  Query Match             7.2%;  Score 213.5;  DB 2;  Length 680;
  Best Local Similarity   21.3%;  Pred. No. 6.5e-06;
  Matches   73;  Conservative   69;  Mismatches  139;  Indels   61;  Gaps    9;

Qy     202 VPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAAEKSITILSTPEGTSAACKSIL 261
           :| : || :||: |  |: |:::::  :: :   | :     :|: |   |   | | |
Db     152 IPEESVGLVIGRNGVEIQAISQKSGCRVQI-VAEPSTTGYRSVDIYGISENIEVAKKLIN 210

Qy     262 EIM---HKEAQD-------------IKFTEEIPLKILAHNNFVGRLIGKEGRNLKKIEQ 304
           |::    | :|:                :  : :: : |    |  | :||||:|  ::|:
Db     211 EVVARGRKLSQEPLPCSVPQFQPIPAVSNSSKVTIIIPIPANKCGAIIGKKGEQMRKLRS 270

Qy     305 DTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIASMNLQAHL 364
             |:     :  :||  : :  : : : |  :   |:                 | | :
Db     271 WTNCDFIL--IQENNIADSVKPLQITGQPKEVEHAKA-----------------LVADI 310

Qy     365 IPGLNLNALGLFPPTSGMPPPTSGPPSAMTPPYPQFEQSETETVHQFIPALSVGAIIGKQ 424
           : | :                ||: |    |    |   :   :|  :||||:| |
Db     311 LDGFD----------------ECPPAGMAGNSPVAAMS----LQVKVPRSTVGAIMGLQ 349

Qy     425 GQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEENFVSPKE 484
           | :||::|      |:  | : |    | ::: |     :    |:  || | |  :
Db     350 GSNIKKISNETETKIQFMPDDDPKLMERTLVVIGNKNKVYVC-ARLLQKIVEANSENANT 408

Qy     485 EVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRD 526
           : |  :: :|:   | |||:||:|: ::   | |   : ||
Db     409 PISL-FYMLIPASKCGLVIGRGGETIRQINKESGAYCEMSRD 449


RESULT 14
S58529
alpha-complex protein 1 - human
N;Alternate names: nucleic acid-binding protein; protein PCBP-1
C;Species: Homo sapiens (man)
C;Date: 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 01-Dec-2000
C;Accession: S58529; S58523; S65678; S43489; S41378; S42472
R;Kiledjian, M.; Wang, X.; Liebhaber, S.A.
EMBO J. 14, 4357-4364, 1995
A;Title: Identification of two KH domain proteins in the alpha-globin mRNP stability
A;Reference number: S58523; MUID:96016208; PMID:7556077
A;Accession: S58529
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-356 <KIL>
A;Accession: S58523
A;Status: preliminary
A;Molecule type: protein
A;Residues: 125-139;251-265;315-323 <KI2>
R;Leffers, H.; Dejgaard, K.; Celis, J.E.
Eur. J. Biochem. 230, 447-453, 1995
A;Title: Characterisation of two major cellular poly(rC)-binding human proteins, each
A;Reference number: S65678; MUID:95331278; PMID:7607214
A;Accession: S65678
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-204,'V',206-356 <LEF>
A;Cross-references: EMBL:X78137; NID:g460770; PIDN:CAA55016.1; PID:g460771
A;Experimental source: AMA cells (transformed human amnion cells)
A;Note: submitted to the EMBL Data Library, March 1994
R;Aasheim, H.C.; Loukianova, T.; Deggerdal, A.; Smeland, E.B.
Nucleic Acids Res. 22, 959-964, 1994
A;Title: Tissue specific expression and cDNA structure of a human transcript encoding
A;Reference number: S43489; MUID:94203810; PMID:8152927
A;Accession: S43489
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 1-298,'H' <AAS>
A;Cross-references: EMBL:Z29505; NID:g444020; PIDN:CAA82631.1; PID:g444021
C;Genetics:
A;Gene: GDB:HNRNPX
A;Cross-references: GDB:344947
C;Keywords: RNA binding

  Query Match             7.1%;  Score 209.5;  DB 2;  Length 356;
```

```
  Best Local Similarity   22.8%;  Pred. No. 4.5e-06;
  Matches   79;  Conservative   70;  Mismatches  123;  Indels   75;  Gaps   12;

Qy    196 LPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAAEKSITILSTPEGTSA 255
          | :|||:  : ||:||||:| ::: | ::: ::|::    : |   : |  |: |  |:|
Db     14 LTIRLLMHGKEVGSIIGKKGESVKRIREESGARINI----SEGNCPERIITLTGP--TNA 67

Qy    256 ACKSILEIMHKEAQDIKFT---------EEIPLKILAHNNFVGRLIGKEGRNLKKIEQDT 306
            |:   |: |  :||  :           : |:::       | |||| |  :|:| : |
Db     68 IFKAFAMIIDKLEEDINSSMTNSTAASRPPVTLRLVVPATQCGSLIGKGGCKIKEIREST 127

Qy    307 DTKITISPLQELTLYNPERTITVKGNVETCAKAEEEI----MKKIRESYENDIASMNLQ- 361
          :: ::   ::   : || ||: |  ::  :  ::|    :: : :| :  : ::  |
Db    128 GAQVQVA--GDMLPNSTERAITIAGVPQSVTECVKQICLVMLETLSQSPQGRVMTIPYQP 185

Qy    362 -AHLIPGLNLNALGLFPPTSGMPPPT---SGPP--------------------------- 390
           | :          :| |  |  |||
Db    186 MPASSPVICAGGQDRCSDAAGYPHATHDLEGPPLDAYSIQGQHTISPLDLAKLNQVARQQ 245

Qy    391 ------------SAMTPPYPQFE------QSETETVHQF-IPALSVGAIIGKQGQHIKQL 431
                     : :    |: :       : |:| |:  ||   :| |||:|| :| ::
Db    246 SHFAMMHGGTGFAGIDSSSPEVKGYWASLDASTQTTHELTIPNNLIGCIIGRQGANINEI 305

Qy    432 SRFAGASIKIA-PAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEE 477
          : :|| |||| | |      | | |||   :  ||  |  ::  |
Db    306 RQMSGAQIKIANPVEGSSG--RQVTITGSAASISLAQYLINARLSSE 350


RESULT 15
A44125
high density lipoprotein-binding protein, 110K - human
C;Species: Homo sapiens (man)
C;Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 24-Sep-1999
C;Accession: A44125
R;McKnight, G.L.; Reasoner, J.; Gilbert, T.; Sundquist, K.O.; Hokland, B.; McKernan, P.A
J. Biol. Chem. 267, 12131-12141, 1992
A;Title: Cloning and expression of a cellular high density lipoprotein-binding protein t
A;Reference number: A44125; MUID:92291094; PMID:1318310
A;Accession: A44125
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1268 <MCK>
A;Cross-references: GB:M64098; GB:M83789; NID:g183891; PIDN:AAA35962.1; PID:g183892
A;Note: sequence extracted from NCBI backbone (NCBIN:106862, NCBIP:106863)
C;Superfamily: vigilin

  Query Match             7.0%;  Score 207.5;  DB 2;  Length 1268;
  Best Local Similarity   23.9%;  Pred. No. 4e-05;
  Matches   95;  Conservative   69;  Mismatches  153;  Indels   81;  Gaps   18;

Qy    202 VPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAAEKSITILSTPEGTSAACKSIL 261
          :|  :   :||| |  ::::  :| :|| : | ::          | |  | ||    |  :|
Db    157 IPKEHHRFVIGKNGEKLQDLELKTATKIQIPRPDD---PSNQIKITGTKEGIEKARHEVL 213

Qy    262 EIMHKEAQDIKFTEEIPLKILAHNNFVG---RLIGKEGRNLKKIEQDTDTKITISPLQEL 318
          |     || :  | : ::   |    |   ||:|          :| |:| |:| | |
Db    214 LI--SAEQDKRAVERLEVEKAFHPFIAGPYNRLVG-------EIMQETGTRINIPP---- 260

Qy    319 TLYNPERT-ITVKGNVETCAKAEEEIMKKIRESYENDIASMNL-----QAHLIPGLNLNA 372
          :  || |  |  |  |:|   | ||| |  :     :: :     |   : |   |:  .
Db    261 --PSVNRTEIVFTGEKEQLAQAVARI-KKIYEEKKKKTTTIAVEVKKSQHKYVIGPKGNS 317

Qy    373 LGLFPPTSG----MPPPTS--------GPP----SAMTPPYPQFEQSETETV------HQ 410
          |     :|    :|| |          | |      |:|  | :     :|        |:  .
Db    318 LQEILERTGVSVEIPPSDSISETVILRGEPEKLGQALTEVYAKANSFTVSSVAAPSWLHR 377

Qy    411 FIPALSVGAIIGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRI 470
          |        ||||:||:: ::::     : |  |  |     : : || |     || :|
Db    378 F--------IIGKKGQNLAKITQ-QMPKVHIEFTEGED----KITLEGPTEDVNVAQEQI 424

Qy    471 YGKIKE----ENFVSPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRD 526
          | :|:      ::|    : |  |:          ||| |  :| ::::     | :|
Db    425 EGMVKDLINRMDYVEINIDHKFHRHL----------IGKSGANINRIKDQYKVSVRIP-- 472

Qy    527 QTPDENDQVVVKITGHFYACQVAQRKIQEILTQVKQHQ 564
            ||      :::| |      | |:|:: |: :::::  :
Db    473 --PDSEKSNLIRIEGDPQGVQQAKRELLELASRMENER 508


Search completed: April 24, 2003, 16:15:58
Job time : 25 secs
```

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: April 16, 2003, 16:30:00 ; Search time 1.33662 Seconds
(without alignments)
1695.712 Million cell updates/sec

Title: US-09-897-778-176_COPY_41_51
Perfect score: 11
Sequence: 1 FVDCPDESWAL 11

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	11	100.0	579	4	O00425	O00425 homo sapien
2	7	63.6	100	11	Q9D054	Q9d054 mus musculu
3	7	63.6	576	13	O42254	O42254 gallus gall
4	7	63.6	577	4	Q9NZI8	Q9nzi8 homo sapien
5	7	63.6	577	11	O88477	O88477 mus musculu
6	7	63.6	579	11	Q9CPN8	Q9cpn8 mus musculu
7	7	63.6	582	13	Q9PW80	Q9pw80 brachydanio
8	7	63.6	593	13	O57526	O57526 xenopus lae
9	7	63.6	594	13	O73932	O73932 xenopus lae
10	6	54.5	225	10	Q8S3U3	Q8s3u3 nicotiana b
11	6	54.5	234	16	Q8UGF6	Q8ugf6 agrobacteri
12	6	54.5	244	16	Q98IF4	Q98if4 rhizobium l
13	6	54.5	250	16	Q8XQB0	Q8xqb0 ralstonia s
14	6	54.5	310	2	P96562	P96562 amycolatops
15	6	54.5	313	17	Q9HP19	Q9hp19 halobacteri
16	6	54.5	337	2	O52816	O52816 amycolatops
17	6	54.5	406	2	O87675	O87675 amycolatops
18	6	54.5	406	2	Q8RN03	Q8rn03 amycolatops
19	6	54.5	460	17	Q9HR19	Q9hr19 halobacteri
20	6	54.5	857	13	P79708	P79708 chiloscylli
21	6	54.5	2338	16	Q8XT19	Q8xt19 ralstonia s
22	5	45.5	88	16	Q9ABN7	Q9abn7 caulobacter
23	5	45.5	101	16	Q8YU34	Q8yu34 anabaena sp
24	5	45.5	104	2	Q9ZND8	Q9znd8 thermus the
25	5	45.5	118	2	O85922	O85922 sphingomona
26	5	45.5	130	16	Q93J47	Q93j47 streptomyce
27	5	45.5	131	10	Q9ATB1	Q9atb1 brassica ol
28	5	45.5	134	10	Q9ATA0	Q9ata0 brassica ca
29	5	45.5	135	10	Q9AXT3	Q9axt3 brassica na
30	5	45.5	136	16	Q9XAE6	Q9xae6 streptomyce
31	5	45.5	140	16	Q9JZU8	Q9jzu8 neisseria m
32	5	45.5	158	10	Q9FR89	Q9fr89 brassica na
33	5	45.5	160	16	Q9JUW2	Q9juw2 neisseria m
34	5	45.5	169	12	P87637	P87637 dhori virus
35	5	45.5	171	10	Q944D7	Q944d7 brassica ol
36	5	45.5	182	17	Q9YDI3	Q9ydi3 aeropyrum p
37	5	45.5	190	16	Q92XI3	Q92xi3 rhizobium m
38	5	45.5	198	16	O66889	O66889 aquifex aeo
39	5	45.5	207	5	Q9U8F4	Q9u8f4 schistosoma
40	5	45.5	209	16	Q8XKF0	Q8xkf0 clostridium
41	5	45.5	211	5	O96303	O96303 trichomonas
42	5	45.5	211	5	O96305	O96305 monocercomo
43	5	45.5	215	5	O96304	O96304 nosema locu
44	5	45.5	215	5	O97437	O97437 giardia lam
45	5	45.5	217	16	Q9KMY8	Q9kmy8 vibrio chol

ALIGNMENTS

RESULT 1
O00425
ID O00425 PRELIMINARY; PRT; 579 AA.
AC O00425;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative RNA binding protein KOC (KOC).
GN KOC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Mueller-Pillasch F., Lacher U., Wallrapp C., Et AL.;
RL Oncogene 0:0-0(0).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=PANCREAS;
RA Mueller-Pillasch F., Lacher U., Wallrapp C.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
DR EMBL; U97188; AAC35208.1; -.
DR EMBL; U76705; AAD09223.1; -.
DR InterPro; IPR004087; KH_dom.
DR InterPro; IPR004088; KH_type_1.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00013; KH-domain; 4.
DR Pfam; PF00076; rrm; 2.
DR SMART; SM00322; KH; 4.
DR SMART; SM00360; RRM; 2.
DR PROSITE; PS50084; KH_TYPE_1; 4.
DR PROSITE; PS50102; RRM; 2.
DR PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
SQ SEQUENCE 579 AA; 63720 MW; AE5C3A8EE3C135C5 CRC64;

Query Match 100.0%; Score 11; DB 4; Length 579;
Best Local Similarity 100.0%; Pred. No. 3.1e-06;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 FVDCPDESWAL 11
           ||||||||||
Db      41 FVDCPDESWAL 51


RESULT 2
Q9D054
ID   Q9D054      PRELIMINARY;      PRT;   100 AA.
AC   Q9D054;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   2610101N11Rik protein.
GN   IGF2BP3 OR 2610101N11RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=EMBRYO;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK011797; BAB27848.1; -.
DR   MGD; MGI:1890359; Igf2bp3.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00076; rrm; 1.
DR   SMART; SM00360; RRM; 1.
DR   PROSITE; PS50102; RRM; 1.
SQ   SEQUENCE   100 AA;  11249 MW;  4D871E37EB9D5466 CRC64;

  Query Match             63.6%;  Score 7;  DB 11;  Length 100;
  Best Local Similarity   100.0%;  Pred. No. 0.11;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 FVDCPDE 7
           |||||||
Db      41 FVDCPDE 47


RESULT 3
O42254
ID   O42254      PRELIMINARY;      PRT;   576 AA.
AC   O42254;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Zipcode-binding protein.
GN   ZBP1.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97220007; PubMed=9121465;
RA   Ross A.F., Oleynikov Y.S., Kislauskis E.H., Taneja K.L., Singer R.H.;
RT   "Characterization of a beta-actin mRNA zipcode-binding protein.";
RL   Mol. Cell. Biol. 17:2158-2165(1997).
DR   EMBL; AF026527; AAB82295.1; -.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00013; KH-domain; 4.
DR   Pfam; PF00076; rrm; 2.
DR   SMART; SM00322; KH; 4.
DR   SMART; SM00360; RRM; 2.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
DR   PROSITE; PS50102; RRM; 2.
DR   PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
SQ   SEQUENCE   576 AA;  63271 MW;  01AAF2D1D81C8811 CRC64;

  Query Match             63.6%;  Score 7;  DB 13;  Length 576;
  Best Local Similarity   100.0%;  Pred. No. 0.43;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 FVDCPDE 7
           |||||||
Db      41 FVDCPDE 47


RESULT 4
Q9NZI8
ID   Q9NZI8      PRELIMINARY;      PRT;   577 AA.
AC   Q9NZI8;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   mRNA-binding protein CRDBP.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ioannidis P., Trangas T., Dimitriadis E., Samiotaki M.,
RA   Panoutsakopoulos G., Kyriazoglou I., Voutzoulias S., Tsiapalis C.M.,
RA   Kittas C., Agnantis N., Pandis N.;
RT   "Ectopic expression of a KH-domain containing protein, highly
RT   homologous to both human IMP-1 and mouse CRD-BP, in benign and
RT   malignant mesenchymal tumors.";
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF198254; AAF37203.1; -.
DR   HSSP; P11940; 1CVJ.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00013; KH-domain; 4.
DR   Pfam; PF00076; rrm; 2.
DR   SMART; SM00322; KH; 4.
DR   SMART; SM00360; RRM; 2.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
DR   PROSITE; PS50102; RRM; 2.
SQ   SEQUENCE   577 AA;  63456 MW;  0749A060F252D81D CRC64;

  Query Match             63.6%;  Score 7;  DB 4;  Length 577;
  Best Local Similarity   100.0%;  Pred. No. 0.43;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 FVDCPDE 7
           |||||||
Db      41 FVDCPDE 47


RESULT 5
```

```
O88477
ID   O88477      PRELIMINARY;      PRT;   577 AA.
AC   O88477;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Coding region determinant binding protein (Coding region determinant-
DE   binding protein).
GN   IGF2BP1 OR CRDBP.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92217743; PubMed=1559612;
RA   Bernstein P.L., Herrick D.J., Prokipcak R.D., Ross J.;
RT   "Control of c-myc mRNA half-life in vitro by a protein capable of
RT   binding to a coding region stability determinant.";
RL   Genes Dev. 6:642-654(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94158886; PubMed=8114742;
RA   Herrick D.J., Ross J.;
RT   "The half-life of c-myc mRNA in growing and serum-stimulated cells:
RT   influence of the coding and 3' untranslated regions and role of
RT   ribosome translocation.";
RL   Mol. Cell. Biol. 14:2119-2128(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94179348; PubMed=8132663;
RA   Prokipcak R.D., Herrick D.J., Ross J.;
RT   "Purification and properties of a protein that binds to the C-terminal
RT   coding region of human c-myc mRNA.";
RL   J. Biol. Chem. 269:9261-9269(1994).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97322234; PubMed=9178888;
RA   Leeds P., Kren B.T., Boylan J.M., Betz N.A., Steer C.J.,
RA   Gruppuso P.A., Ross J.;
RT   "Developmental regulation of CRD-BP, an RNA-binding protein that
RT   stabilizes c-myc mRNA in vitro.";
RL   Oncogene 14:1279-1286(1997).
RN   [5]
RP   SEQUENCE FROM N.A.
RA   Ross J., Prokipcak R.D., Leeds P., Doyle G.A.R., Betz N.A.,
RA   Fleisig A.J.;
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=EMBRYONIC HEAD;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AF061569; AAC72743.1; -.
DR   EMBL; AK013940; BAB29071.1; -.
DR   HSSP; P11940; 1CVJ.
DR   MGD; MGI:1890357; Igf2bp1.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00013; KH-domain; 4.
DR   Pfam; PF00076; rrm; 2.
DR   SMART; SM00322; KH; 4.
DR   SMART; SM00360; RRM; 2.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
DR   PROSITE; PS50102; RRM; 2.
SQ   SEQUENCE   577 AA;  63450 MW;  EFBB1AF2FF9F0344 CRC64;

  Query Match           63.6%;  Score 7;  DB 11;  Length 577;
  Best Local Similarity  100.0%;  Pred. No. 0.43;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 FVDCPDE 7
           |||||||
Db      41 FVDCPDE 47


RESULT 6
Q9CPN8
ID   Q9CPN8      PRELIMINARY;      PRT;   579 AA.
AC   Q9CPN8;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   10 days embryo cDNA, RIKEN full-length enriched library,
DE   clone:2610036B18, full insert sequence (Igf2 mRNA-binding protein
DE   3).
GN   IGF2BP3 OR 2610101N11RIK OR MIMP3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=EMBRYO;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Mori H., Sakakibara S., Imai T., Nakamura Y., Iijima T., Suzuki A.,
RA   Yuasa Y., Takeda M., Okano H.;
RT   "Expression of mouse igf2 mRNA-binding protein 3 and its implications
RT   for the developing central nervous system.";
RL   J. Neurosci. Res. 0:0-0(2001).
DR   EMBL; AK011689; BAB27779.1; -.
DR   EMBL; AB046173; BAB19755.1; -.
DR   MGD; MGI:1890359; Igf2bp3.
DR   InterPro; IPR004087; KH_dom.
```

```
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00013; KH-domain; 4.
DR   Pfam; PF00076; rrm; 2.
DR   SMART; SM00322; KH; 4.
DR   SMART; SM00360; RRM; 2.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
DR   PROSITE; PS50102; RRM; 2.
SQ   SEQUENCE   579 AA;  63574 MW;  CABD9A4355B392B7 CRC64;

  Query Match           63.6%;  Score 7;  DB 11;  Length 579;
  Best Local Similarity  100.0%;  Pred. No. 0.44;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 FVDCPDE 7
          |||||||
Db     41 FVDCPDE 47


RESULT 7
Q9PW80
ID   Q9PW80      PRELIMINARY;      PRT;   582 AA.
AC   Q9PW80;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Vg1 RNA binding protein.
GN   DVR1RBP.
OS   Brachydanio rerio (Zebrafish) (Zebra danio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Zhang Q., Yaniv K., Oberman F., Wolke U., Git A., Fromer M.,
RA   Taylor W., Meyer D., Standart N., Raz E., Yisraeli J.K.;
RT   "Vg1 RBP intracellular distribution and evolutionarily conserved
RT   expression suggest multiple roles during development.";
RL   Mech. Dev. 0:0-0(1999).
DR   EMBL; AF161270; AAD45610.1; -.
DR   ZFIN; ZDB-GENE-000308-1; dvr1rbp.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00013; KH-domain; 4.
DR   Pfam; PF00076; rrm; 2.
DR   SMART; SM00322; KH; 4.
DR   SMART; SM00360; RRM; 2.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
DR   PROSITE; PS50102; RRM; 2.
SQ   SEQUENCE   582 AA;  63351 MW;  9DAE63200681B306 CRC64;

  Query Match           63.6%;  Score 7;  DB 13;  Length 582;
  Best Local Similarity  100.0%;  Pred. No. 0.44;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 FVDCPDE 7
          |||||||
Db     41 FVDCPDE 47


RESULT 8
O57526
ID   O57526      PRELIMINARY;      PRT;   593 AA.
AC   O57526;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   KH domain-containing transcription factor B3.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92249652; PubMed=1577195;
RA   Pfaff S.L., Taylor W.L.;
RT   "Characterization of a Xenopus oocyte factor that binds to a
RT   developmentally regulated cis-element in the TFIIIA gene.";
RL   Dev. Biol. 151:306-316(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Griffin D., Taylor W.L.;
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Havin L., Git A., Elisha Z., Oberman F., Yaniv K.,
RA   Pressman Schwartz S., Standart N.M., Yisraeli J.K.;
RL   Genes Dev. 0:0-0(1998).
DR   EMBL; AF042353; AAB97457.1; -.
DR   EMBL; AF064633; AAC18597.1; -.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00013; KH-domain; 4.
DR   Pfam; PF00076; rrm; 2.
DR   SMART; SM00322; KH; 4.
DR   SMART; SM00360; RRM; 2.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
DR   PROSITE; PS50102; RRM; 2.
DR   PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
SQ   SEQUENCE   593 AA;  65385 MW;  5A5AB4B4A1D55DF7 CRC64;

  Query Match           63.6%;  Score 7;  DB 13;  Length 593;
  Best Local Similarity  100.0%;  Pred. No. 0.44;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 FVDCPDE 7
          |||||||
Db     41 FVDCPDE 47


RESULT 9
O73932
ID   O73932      PRELIMINARY;      PRT;   594 AA.
AC   O73932;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   VG1 RNA binding protein variant D.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Havin L., Git A., Elisha Z., Oberman F., Yaniv K.,
RA   Pressman Schwartz S., Standart N.M., Yisraeli J.K.;
RL   Genes Dev. 0:0-0(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98228351; PubMed=9560341;
RA   Deshler J.O., Highett M.I., Abramson T., Schnapp B.J.;
RT   "A highly conserved RNA-binding protein for cytoplasmic mRNA
RT   localization in vertebrates.";
RL   Curr. Biol. 8:489-496(1998).
DR   EMBL; AF064634; AAC18598.1; -.
DR   EMBL; AF055923; AAC41285.1; -.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00013; KH-domain; 4.
DR   Pfam; PF00076; rrm; 2.
```

```
DR   SMART; SM00322; KH; 4.
DR   SMART; SM00360; RRM; 2.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
DR   PROSITE; PS50102; RRM; 2.
SQ   SEQUENCE   594 AA;  65643 MW;  54CEA7BFF0856DD6 CRC64;

  Query Match           63.6%;  Score 7;  DB 13;  Length 594;
  Best Local Similarity  100.0%;  Pred. No. 0.44;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 FVDCPDE 7
          |||||||
Db     41 FVDCPDE 47


RESULT 10
Q8S3U3
ID   Q8S3U3      PRELIMINARY;      PRT;   225 AA.
AC   Q8S3U3;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Prf-like protein (Fragment).
OS   Nicotiana benthamiana.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
OX   NCBI_TaxID=4100;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21864904; PubMed=11874570;
RA   Peart J.R., Cook G., Feys B.J., Parker J.E., Baulcombe D.C.;
RT   "An EDS1 orthologue is required for N-mediated resistance against
RT   tobacco mosaic virus.";
RL   Plant J. 29:569-579(2002).
DR   EMBL; AF479624; AAL85346.1; -.
FT   NON_TER       1      1
FT   NON_TER     225    225
SQ   SEQUENCE   225 AA;  25674 MW;  1E173D5E15F9860B CRC64;

  Query Match           54.5%;  Score 6;  DB 10;  Length 225;
  Best Local Similarity  100.0%;  Pred. No. 4;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6 DESWAL 11
          ||||||
Db    134 DESWAL 139


RESULT 11
Q8UGF6
ID   Q8UGF6      PRELIMINARY;      PRT;   234 AA.
AC   Q8UGF6;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Branched-chain amino acid permease.
GN   ATU1082 OR AGR_C_2001.
OS   Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Rhizobiaceae; Rhizobium.
OX   NCBI_TaxID=176299;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21608550; PubMed=11743193;
RA   Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA   Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA   Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA   Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA   Kutyavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA   Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA   Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA   Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA   Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA   Nester E.W.;
RT   "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT   C58.";
RL   Science 294:2317-2323(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21608551; PubMed=11743194;
RA   Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA   Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA   Houmiel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA   Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA   Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,
RA   Cielo C., Slater S.;
RT   "Genome sequence of the plant pathogen and biotechnology agent
RT   Agrobacterium tumefaciens C58.";
RL   Science 294:2323-2328(2001).
DR   EMBL; AE009072; AAL42095.1; -.
DR   EMBL; AE008038; AAK86891.1; -.
KW   Complete proteome.
SQ   SEQUENCE   234 AA;  24282 MW;  2D6931DA08A13758 CRC64;

  Query Match           54.5%;  Score 6;  DB 16;  Length 234;
  Best Local Similarity  100.0%;  Pred. No. 4.1;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6 DESWAL 11
          ||||||
Db    113 DESWAL 118


RESULT 12
Q98IF4
ID   Q98IF4      PRELIMINARY;      PRT;   244 AA.
AC   Q98IF4;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein mll2427.
GN   MLL2427.
OS   Rhizobium loti (Mesorhizobium loti).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Phyllobacteriaceae; Mesorhizobium.
OX   NCBI_TaxID=381;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MAFF303099;
RX   MEDLINE=21082930; PubMed=11214968;
RA   Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA   Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA   Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA   Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA   Takeuchi C., Yamada M., Tabata S.;
RT   "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT   Mesorhizobium loti.";
RL   DNA Res. 7:331-338(2000).
DR   EMBL; AP002999; BAB49562.1; -.
DR   InterPro; IPR004471; AzlC.
DR   Pfam; PF03591; AzlC; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   244 AA;  25766 MW;  E0DCF859FF3CA77F CRC64;

  Query Match           54.5%;  Score 6;  DB 16;  Length 244;
  Best Local Similarity  100.0%;  Pred. No. 4.2;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6 DESWAL 11
          ||||||
Db    121 DESWAL 126


RESULT 13
Q8XQB0
```

```
ID   Q8XQB0      PRELIMINARY;      PRT;   250 AA.
AC   Q8XQB0;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical transmembrane protein RSp1376.
GN   RSP1376 OR RS02082.
OS   Ralstonia solanacearum (Pseudomonas solanacearum).
OG   Plasmid megaplasmid.
OC   Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC   Ralstonia.
OX   NCBI_TaxID=305;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GMI1000;
RX   MEDLINE=21681879; PubMed=11823852;
RA   Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA   Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA   Chandler M., Choisne N., Claudel-Renard C., Cunnac S., Demange N.,
RA   Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA   Siguier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA   Weissenbach J., Boucher C.A.;
RT   "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL   Nature 415:497-502(2002).
DR   EMBL; AL646084; CAD18527.1; -.
DR   InterPro; IPR004471; AzlC.
DR   Pfam; PF03591; AzlC; 1.
KW   Plasmid; Hypothetical protein; Complete proteome.
SQ   SEQUENCE   250 AA;  26490 MW;  685F6FB89089EEC2 CRC64;

  Query Match             54.5%;  Score 6;  DB 16;  Length 250;
  Best Local Similarity   100.0%;  Pred. No. 4.3;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 DESWAL 11
           ||||||
Db     127 DESWAL 132


RESULT 14
P96562
ID   P96562      PRELIMINARY;      PRT;   310 AA.
AC   P96562;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Putative cytochrome P450 165C2 (Hypothetical hydroxylase D)
DE   (Fragment).
GN   CYP165C2.
OS   Amycolatopsis orientalis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Pseudonocardineae; Pseudonocardiaceae; Amycolatopsis.
OX   NCBI_TaxID=31958;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C329.4;
RA   Solenberg P.J., Matsushima P., Stack D.R., Wilkie S.C., Thompson R.C.,
RA   Baltz R.H.;
RL   Submitted (MAR-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR   EMBL; U84350; AAB49296.1; -.
DR   HSSP; Q00441; 1OXA.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   Pfam; PF00067; p450; 1.
DR   PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW   Electron transport; Heme; Hypothetical protein; Membrane;
KW   Monooxygenase; Oxidoreductase.
FT   NON_TER       1      1
FT   BINDING     260    260       HEME (BY SIMILARITY).
SQ   SEQUENCE   310 AA;  34255 MW;  0155572AFA694487 CRC64;

  Query Match             54.5%;  Score 6;  DB 2;  Length 310;
  Best Local Similarity   100.0%;  Pred. No. 5.1;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 FVDCPD 6
           ||||||
Db     299 FVDCPD 304


RESULT 15
Q9HP19
ID   Q9HP19      PRELIMINARY;      PRT;   313 AA.
AC   Q9HP19;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Vng1845c.
GN   VNG1845C.
OS   Halobacterium sp. (strain NRC-1).
OC   Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC   Halobacteriaceae; Halobacterium.
OX   NCBI_TaxID=64091;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20504483; PubMed=11016950;
RA   Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA   Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA   Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA   Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA   Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA   Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA   Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA   Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
RT   "Genome sequence of Halobacterium species NRC-1.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR   EMBL; AE005085; AAG20051.1; -.
DR   InterPro; IPR000182; GCN5acetyltransf.
DR   Pfam; PF00583; Acetyltransf; 1.
KW   Complete proteome.
SQ   SEQUENCE   313 AA;  34330 MW;  6233C497ED4F8303 CRC64;

  Query Match             54.5%;  Score 6;  DB 17;  Length 313;
  Best Local Similarity   100.0%;  Pred. No. 5.2;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 DESWAL 11
           ||||||
Db     186 DESWAL 191


Search completed: April 16, 2003, 16:36:41
Job time : 3.33662 secs
```

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: April 24, 2003, 16:11:07 ; Search time 38 Seconds
(without alignments)
3139.507 Million cell updates/sec

Title: US-09-897-778-176
Perfect score: 2956
Sequence: 1 MNKLYIGNLSENAAPSDLES.....VKQHQQQKALQSGPPQSRRK 579

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2956	100.0	579	4	O00425	O00425 homo sapien
2	2858	96.7	579	11	Q9CPN8	Q9cpn8 mus musculu
3	2483	84.0	593	13	O57526	O57526 xenopus lae
4	2461.5	83.3	594	13	O73932	O73932 xenopus lae
5	2261.5	76.5	582	13	Q9PW80	Q9pw80 brachydanio
6	2219.5	75.1	576	13	O42254	O42254 gallus gall
7	2192	74.2	577	11	O88477	O88477 mus musculu
8	2181	73.8	577	4	Q9NZI8	Q9nzi8 homo sapien
9	1868	63.2	556	4	Q9Y6M1	Q9y6m1 homo sapien
10	835.5	28.3	558	5	Q9VZ69	Q9vz69 drosophila
11	835	28.2	566	5	Q9NGS9	Q9ngs9 drosophila
12	473	16.0	100	11	Q9D054	Q9d054 mus musculu
13	418.5	14.2	854	5	Q21605	Q21605 caenorhabdi
14	262.5	8.9	774	10	Q9LI28	Q9li28 oryza sativ
15	260.5	8.8	542	10	Q8S7G1	Q8s7g1 oryza sativ
16	242.5	8.2	557	5	Q23487	Q23487 caenorhabdi
17	240	8.1	644	4	Q12828	Q12828 homo sapien
18	240	8.1	651	11	Q91WJ8	Q91wj8 mus musculu
19	238	8.1	641	5	Q9BLA0	Q9bla0 caenorhabdi
20	236.5	8.0	568	10	Q9LXF5	Q9lxf5 arabidopsis
21	235.5	8.0	621	10	Q9C553	Q9c553 arabidopsis
22	235.5	8.0	653	4	Q96AE4	Q96ae4 homo sapien
23	234	7.9	600	4	Q92946	Q92946 homo sapien
24	233	7.9	492	4	Q9UNW9	Q9unw9 homo sapien
25	233	7.9	498	4	O43267	O43267 homo sapien
26	231.5	7.8	313	10	Q9LZ82	Q9lz82 arabidopsis
27	231	7.8	364	5	Q95SZ9	Q95sz9 drosophila
28	230.5	7.8	493	11	Q9JKN6	Q9jkn6 mus musculu
29	230	7.8	618	5	Q9BLA1	Q9bla1 caenorhabdi
30	229.5	7.8	398	3	O74919	O74919 schizosacch
31	229	7.7	386	5	P91632	P91632 drosophila
32	228.5	7.7	313	4	Q96EP6	Q96ep6 homo sapien
33	226	7.6	1301	5	Q9U982	Q9u982 drosophila
34	226	7.6	1301	5	Q9V8H6	Q9v8h6 drosophila
35	225.5	7.6	577	10	Q9SR13	Q9sr13 arabidopsis
36	225.5	7.6	589	5	Q17935	Q17935 caenorhabdi
37	225.5	7.6	611	5	Q17936	Q17936 caenorhabdi
38	224	7.6	769	13	Q8UVD9	Q8uvd9 gallus gall
39	222	7.5	479	10	Q9XI71	Q9xi71 arabidopsis
40	219.5	7.4	396	13	Q9PSI1	Q9psi1 xenopus lae
41	217.5	7.4	640	10	Q9ASX3	Q9asx3 arabidopsis
42	217.5	7.4	644	10	Q9FNK3	Q9fnk3 arabidopsis
43	215.5	7.3	403	11	Q9DB01	Q9db01 mus musculu
44	214	7.2	833	10	Q9FMF0	Q9fmf0 arabidopsis
45	213.5	7.2	510	10	Q8W5C2	Q8w5c2 oryza sativ

ALIGNMENTS

RESULT 1
O00425
ID O00425 PRELIMINARY; PRT; 579 AA.
AC O00425;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative RNA binding protein KOC (KOC).
GN KOC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Mueller-Pillasch F., Lacher U., Wallrapp C., Et AL.;
RL Oncogene 0:0-0(0).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=PANCREAS;
RA Mueller-Pillasch F., Lacher U., Wallrapp C.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
DR EMBL; U97188; AAC35208.1; -.
DR EMBL; U76705; AAD09223.1; -.
DR InterPro; IPR004087; KH_dom.
DR InterPro; IPR004088; KH_type_1.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00013; KH-domain; 4.
DR Pfam; PF00076; rrm; 2.
DR SMART; SM00322; KH; 4.
DR SMART; SM00360; RRM; 2.
DR PROSITE; PS50084; KH_TYPE_1; 4.
DR PROSITE; PS50102; RRM; 2.
DR PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
SQ SEQUENCE 579 AA; 63720 MW; AE5C3A8EE3C135C5 CRC64;

Query Match 100.0%; Score 2956; DB 4; Length 579;
Best Local Similarity 100.0%; Pred. No. 1.7e-187;
Matches 579; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGK 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGK 60

Qy      61 IELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 IELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETA 120

Qy     121 VVNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDEMAAQQNPLQQPRGRRGLGQRGSS 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 VVNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDEMAAQQNPLQQPRGRRGLGQRGSS 180

Qy     181 RQGSPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAA 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 RQGSPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAA 240

Qy     241 EKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEGRNLK 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 EKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEGRNLK 300

Qy     301 KIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIASMNL 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 KIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIASMNL 360

Qy     361 QAHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMTPPYPQPEQSETETVHQFIPALSVGAI 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 QAHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMTPPYPQPEQSETETVHQFIPALSVGAI 420

Qy     421 IGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEENFV 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 IGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEENFV 480

Qy     481 SPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVVVKIT 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 SPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVVVKIT 540

Qy     541 GHFYACQVAQRKIQEILTQVKQHQQQKALQSGPPQSRRK 579
           |||||||||||||||||||||||||||||||||||||||
Db     541 GHFYACQVAQRKIQEILTQVKQHQQQKALQSGPPQSRRK 579
```

```
RESULT 2
Q9CPN8
ID   Q9CPN8      PRELIMINARY;      PRT;   579 AA.
AC   Q9CPN8;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   10 days embryo cDNA, RIKEN full-length enriched library,
DE   clone:2610036B18, full insert sequence (Igf2 mRNA-binding protein
DE   3).
GN   IGF2BP3 OR 2610101N11RIK OR MIMP3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=EMBRYO;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Mori H., Sakakibara S., Imai T., Nakamura Y., Iijima T., Suzuki A.,
RA   Yuasa Y., Takeda M., Okano H.;
RT   "Expression of mouse igf2 mRNA-binding protein 3 and its implications
RT   for the developing central nervous system.";
RL   J. Neurosci. Res. 0:0-0(2001).
DR   EMBL; AK011689; BAB27779.1; -.
DR   EMBL; AB046173; BAB19755.1; -.
DR   MGD; MGI:1890359; Igf2bp3.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00013; KH-domain; 4.
DR   Pfam; PF00076; rrm; 2.
DR   SMART; SM00322; KH; 4.
DR   SMART; SM00360; RRM; 2.
DR   PROSITE; PS50084; KH_TYPE_1; 4.
DR   PROSITE; PS50102; RRM; 2.
SQ   SEQUENCE   579 AA;  63574 MW;  CABD9A4355B392B7 CRC64;
```

```
  Query Match           96.7%;  Score 2858;  DB 11;  Length 579;
  Best Local Similarity  96.4%;  Pred. No. 5.2e-181;
  Matches  558;  Conservative   10;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       1 MNKLYIGNLSENAAPSDLESIFKDAKIPVSGPFLVKTGYAFVDCPDESWALKAIEALSGK 60
           ||||||||||::| |:||||:||||||||:||||||||||||||||| ||||||||||||
Db       1 MNKLYIGNLSDHAGPADLESVFKDAKIPVAGPFLVKTGYAFVDCPDEGWALKAIEALSGK 60

Qy      61 IELHGKPIEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETA 120
           :||||||:||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 MELHGKPMEVEHSVPKRQRIRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETA 120

Qy     121 VVNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDEMAAQQNPLQQPRGRRGLGQRGSS 180
           |||||||||||||||||||||||||||||||||| ||||||  | ||||| ||||||||||
Db     121 VVNVTYSSKDQARQALDKLNGFQLENFTLKVAYIPDETAAQQNPSPQLRGRRGPGQRGSS 180

Qy     181 RQGSPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAA 240
           || ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db     181 RQASPGSVSKQKPCDLPLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENTGAA 240

Qy     241 EKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEGRNLK 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 EKSITILSTPEGTSAACKSILEIMHKEAQDIKFTEEIPLKILAHNNFVGRLIGKEGRNLK 300

Qy     301 KIEQDTDTKITISPLQELTLYNPERTITVKGNVETCAKAEEEIMKKIRESYENDIASMNL 360
           ||||||||||||||||||||||||||||||||:||||||||||||||||||||||||||||
Db     301 KIEQDTDTKITISPLQELTLYNPERTITVKGSVETCAKAEEEIMKKIRESYENDIASMNL 360

Qy     361 QAHLIPGLNLNALGLFPPTSGMPPPTSGPPSAMTPPYPQPEQSETETVHQFIPALSVGAI 420
           |||||||||||||||||||||||||||||| :|||||||||||||| ||||||||||
Db     361 QAHLIPGLNLNALGLFPPTSGMPPPTSGPPSTLTPPYPQPEQSETETVHLFIPALSVGAI 420

Qy     421 IGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEENFV 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 IGKQGQHIKQLSRFAGASIKIAPAEAPDAKVRMVIITGPPEAQFKAQGRIYGKIKEENFV 480

Qy     481 SPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQNLSSAEVVVPRDQTPDENDQVVVKIT 540
           ||||||||||||||||||||||||||||||||||||:||||||||||||||||||||||||
Db     481 SPKEEVKLEAHIRVPSFAAGRVIGKGGKTVNELQSLSSAEVVVPRDQTPDENDQVVVKIT 540

Qy     541 GHFYACQVAQRKIQEILTQVKQHQQQKALQSGPPQSRRK 579
           |||||||||||||||||||||||||||||||||||||||
Db     541 GHFYACQVAQRKIQEILTQVKQHQQQKALQSGPPQSRRK 579
```

RESULT 3
Q57526
ID
AC
OS
24 16:2